Medical Imaging with Deep Learning Overview

Chapter 1 - Radiology and Multi-View

Chapter 2 - Histology and Segmentation

Chapter 3 - Cell Counting

Chapter 4 - Incorrect Feature Attribution

Chapter 5 - GANs in Medical Imaging

Slides by Joseph Paul Cohen 2020
Email: joseph@josephpcohen.com
License: Creative Commons Attribution-Sharealike

Chapter 1

Radiology and Multi-View

Common X-ray projections/views

Most common

(a) P-A

(b) Lateral

(e) A-P

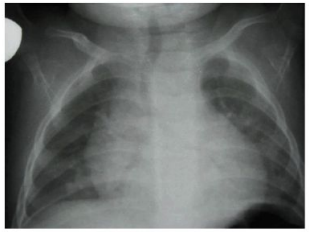
(c) Lordotic

(d) A-P supine

R L

(f) P-A

(g) Lateral



(j) A-P

(h) Lordotic

(i) A-P supine

PA = PosteroAnterior = BackFront

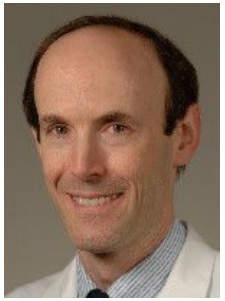
Image: [Bustos, “PadChest: A Large Chest x-Ray Image Dataset with Multi-Label Annotated Reports.” 2019]

Chest X-ray14 Dataset

Released 2017, first large scale chest X-ray dataset

>100k frontal images released as public domain

Enabled the deep learning radiology revolution



Ronald Summers
NIH Clinical Center

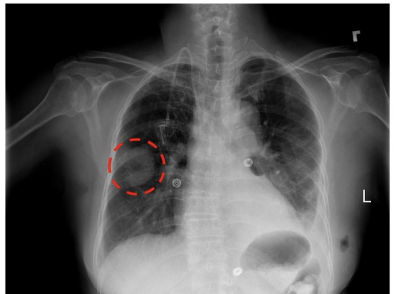
Media Advisory Wednesday, September 27, 2017

NIH Clinical Center provides one of the largest publicly available chest x-ray datasets to scientific community

The dataset of scans is from more than 30,000 patients, including many with advanced lung disease.

What

The NIH Clinical Center recently released over 100,000 anonymized chest x-ray images and their corresponding data to the scientific community. The release will allow researchers across the country and around the world to freely access the datasets and increase their ability to teach computers how to detect and diagnose disease. Ultimately, this artificial intelligence mechanism can lead to clinicians making better diagnostic decisions for patients.



NIH compiled the dataset of scans from more than 30,000 patients, including

Stanford Pneumonia study

In 2017 Pranav Rajpurkar and Jeremy Irvin trained a DenseNet on NIH data scaled to 224x224 pixels

Set the benchmark performance which has not been significantly improved.

They evaluated pneumonia predictions against 4 radiologists.

"We find that the model exceeds the average radiologist performance on the pneumonia detection task."

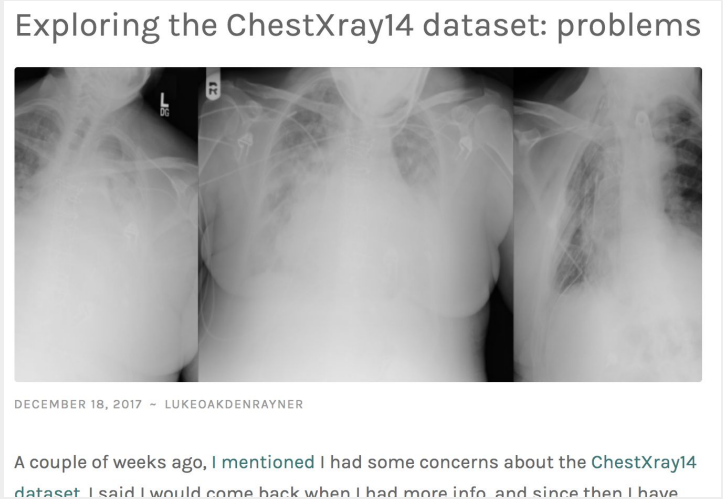
	F1 Score (95% CI)
Radiologist 1	0.383 (0.309, 0.453)
Radiologist 2	0.356 (0.282, 0.428)
Radiologist 3	0.365 (0.291, 0.435)
Radiologist 4	0.442 (0.390, 0.492)
Radiologist Avg.	0.387 (0.330, 0.442)
CheXNet	0.435 (0.387, 0.481)

https://stanfordmlgroup.github.io/projects/chexnet/

Criticism of the Chest X-ray14 Dataset

In 2017 Luke Oakden-Rayner published a blog post discussing issues with the labels in the NIH data.

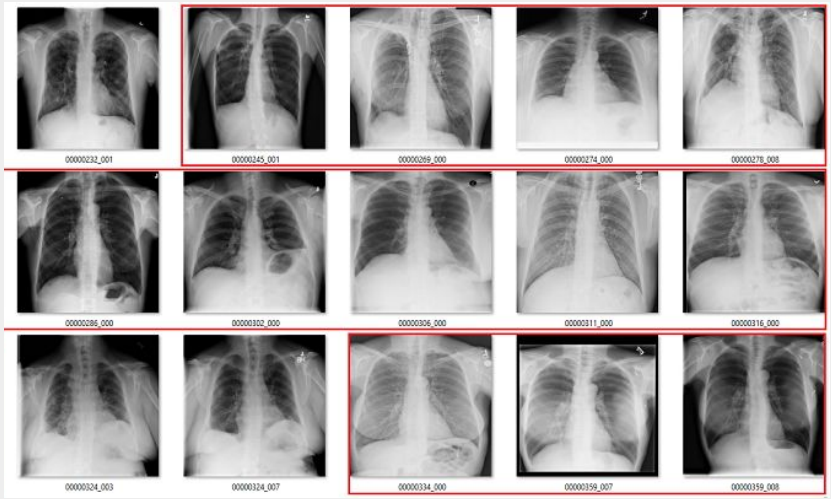
This led to more work on automatic label extraction.


Exploring the ChestXray14 dataset: problems

DECEMBER 18, 2017 ~ LUKEOAKDENRAYNER

A couple of weeks ago, I mentioned I had some concerns about the ChestXray14 dataset. I said I would come back when I had more info, and since then I have

https://lukeoakdenrayner.wordpress.com/2017/12/18/the-chestxray14-dataset-problems/



In a sample of images red are said to be wrong

2019: the year of chest X-ray data

PADCHEST
160k images
Multiple views
Almost 200 labels

27% hand labelled, others using an RNN.

License:Creative Commons Attribution-ShareAlike

CheXpert

CheXpert
224k images
PA and L views
13 labels.

Automated rule-based labeler

Non-commercial research purposes only

MIMIC-CXR
377k images
PA and L views
13 labels.

Automated rule-based labeler. NIH (NegBio) and CheX labelers ran.

Non-commercial research purposes only. Confidentially training required.

Many datasets exist with different methods of obtaining labels. Automatic or hand labelled

NIH chest X-ray14
14 labels

Automated rule-based labeler (NegBio)

PADCHEST, ~200 labels

27% hand labelled, others using an RNN.

CheXpert, 13 labels

Custom rule-based labeler.

MIMIC-CXR, 13 labels

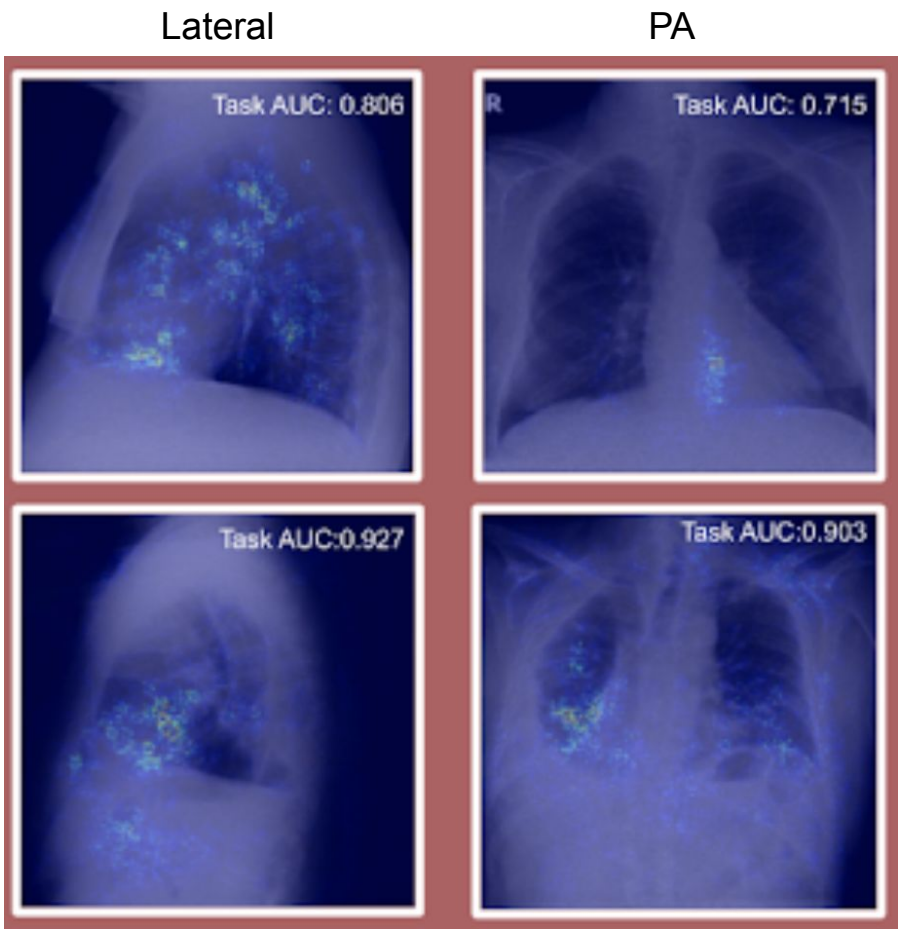
Automated rule-based labeler. NIH (NegBio) and CheX labelers used.

RSNA Pneumonia Kaggle
Relabelled NIH data

A group at Google relabelled a subset of NIH images

MeSH automatic labeller

Multi-modal/view inference (X-ray use case)

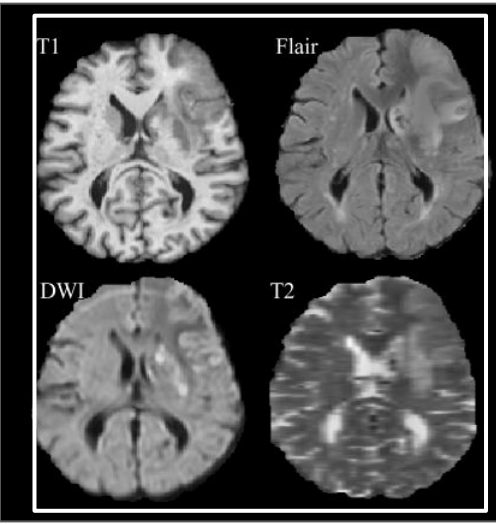


Here saliency maps are from models trained on single views.

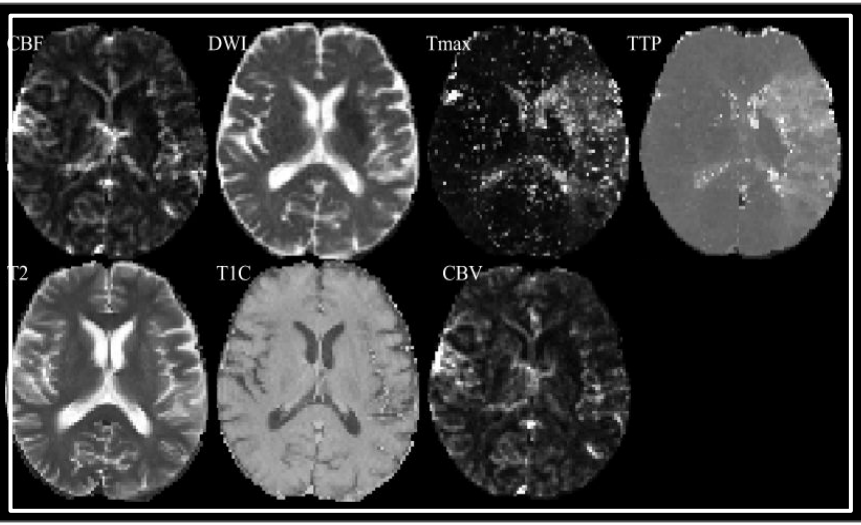
These two tasks perform better when using lateral views.

[Bertrand, 2019]

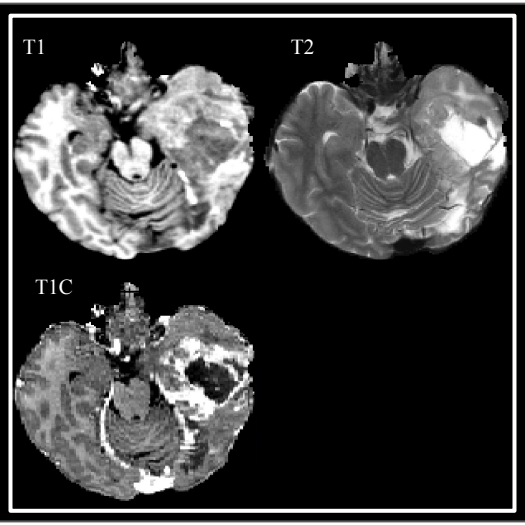
Also: Multi-modal/view inference (MRI use case)



Ischemic stroke lesion segmentation (ISLES dataset)

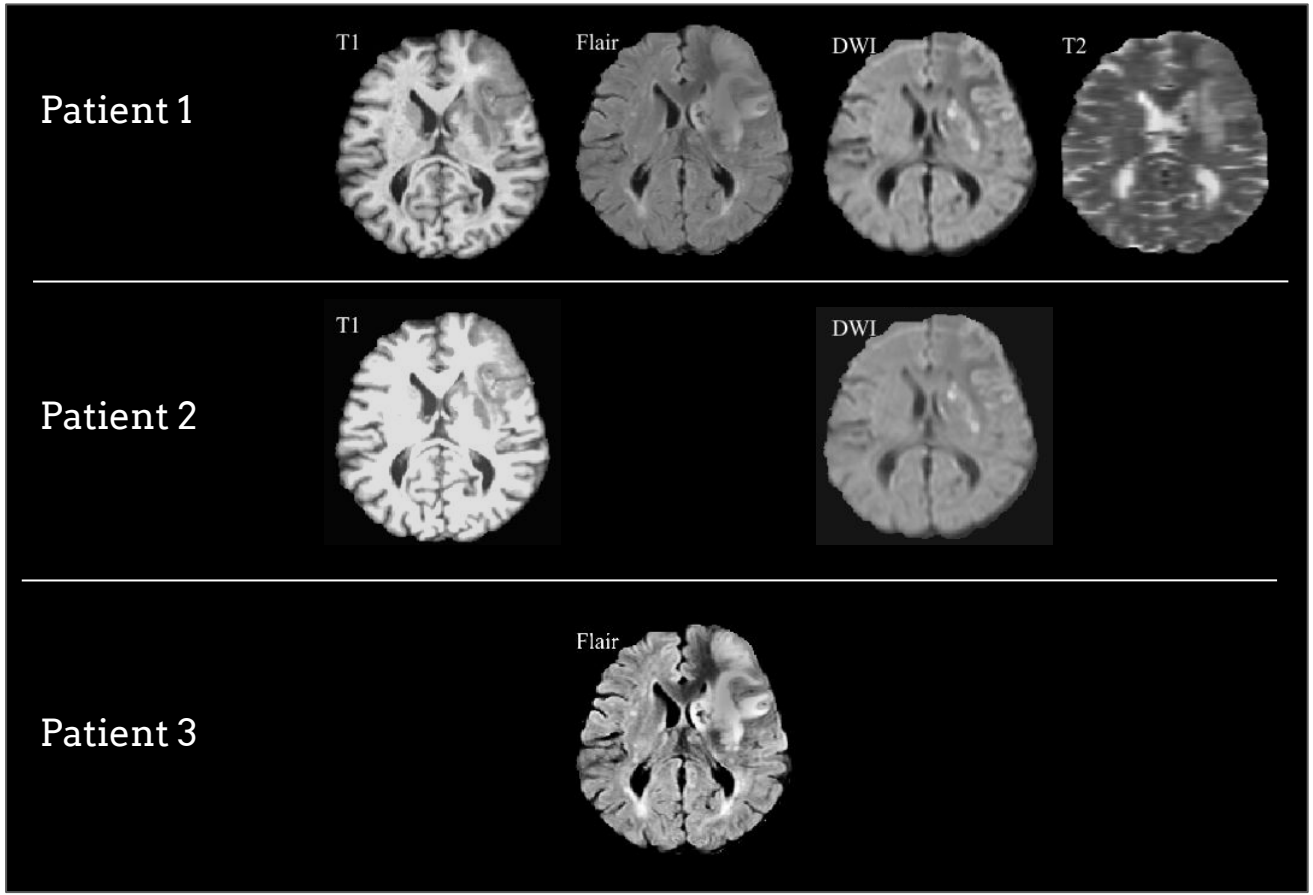


Stroke perfusion estimation



Brain tumor segmentation (BraTS dataset)

Challenge: missing modalities/views



Incomplete Input!

Expected:

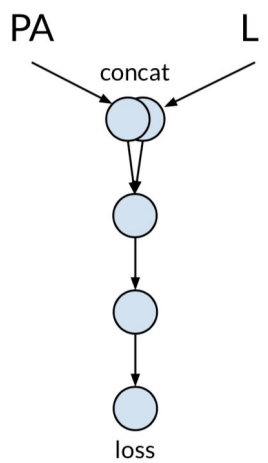
$$f(\text{T1, Flair, DWI, T2})$$

Given:

$$f(\text{T1, - , DWI, - })$$

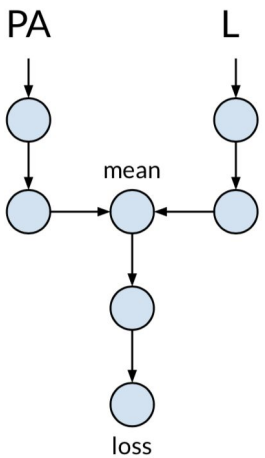
$$f(\text{ - , Flair, - , -})$$

Integrating multiple views



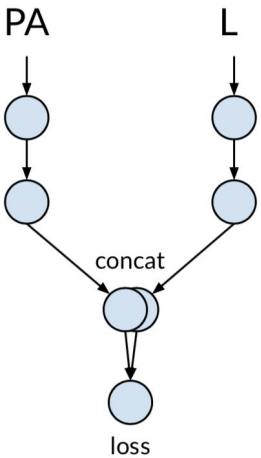
Stacked

Combine images right at the input



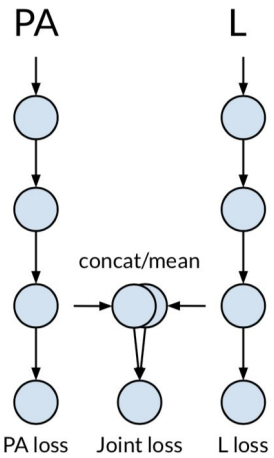
HeMIS

Take mean of activations in the middle of the network



DualNet

Concat output features of two models with single prediction

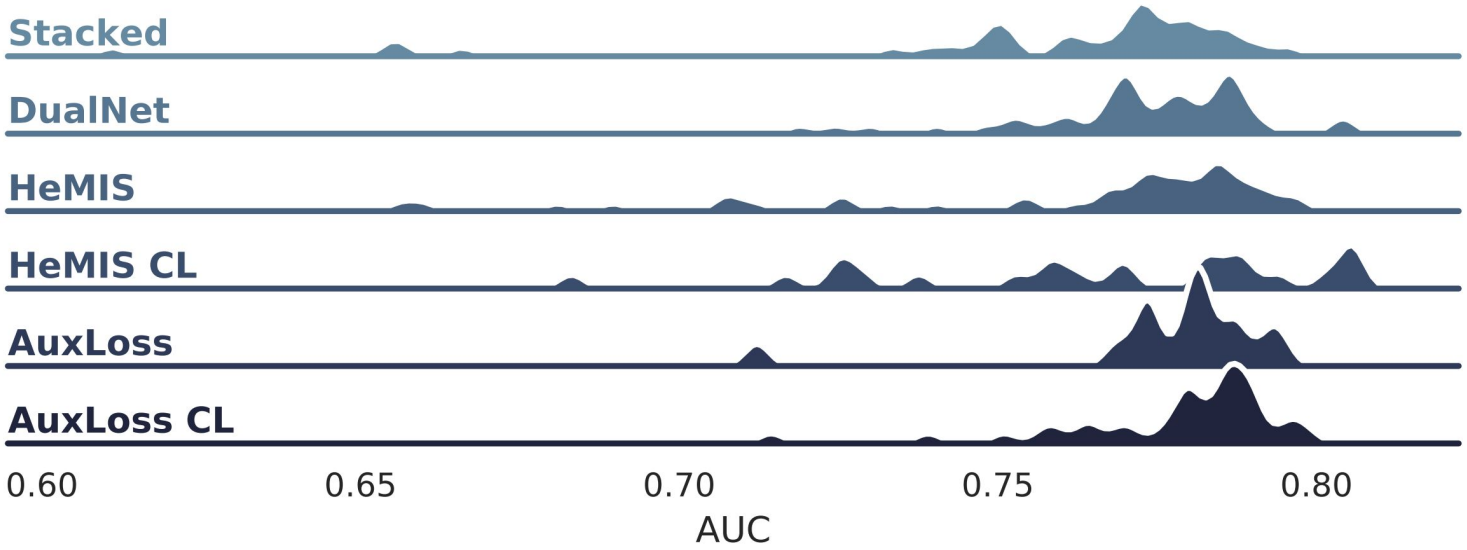


AuxLoss

Three losses. A network for each modality with losses that regularize each network.

Image: [Hashir, Quantifying the Value of Lateral Views in Deep Learning for Chest X-rays, 2020]

Integrating multiple views (X-ray images)



All models are about equal in performance given the right hyperparameters.
Hyperparameter tuning is easier on some models but not others

Image: [Hashir, Quantifying the Value of Lateral Views in Deep Learning for Chest X-rays, 2020]

Chapter 1 - References

Bustos, A., Pertusa, A., Salinas, J.-M., & de la Iglesia-Vayá, M. (2019). PadChest: A large chest x-ray image dataset with multi-label annotated reports. ArXiv Preprint. http://arxiv.org/abs/1901.07441

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR.2017.369

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *Arxiv*. http://arxiv.org/abs/1711.05225

Viviano, J. D., Simpson, B., Dutil, F., Bengio, Y., & Cohen, J. P. (2019). Underwhelming Generalization Improvements From Controlling Feature Attribution. *Arxiv:1910.00199*. http://arxiv.org/abs/1910.00199

Hashir, M., Bertrand, H., & Cohen, J. P. (2020). Quantifying the Value of Lateral Views in Deep Learning for Chest X-rays. *Medical Imaging with Deep Learning*.

Havaei, M., Guizard, N., Chapados, N., & Bengio, Y. (2016). HeMIS: Hetero-modal image segmentation. Medical Image Computing and Computer Assisted Intervention, 9901 LNCS. https://doi.org/10.1007/978-3-319-46723-8_54

Rubin, J., Sanghavi, D., Zhao, C., Lee, K., Qadir, A., & Xu-Wilson, M. (2018, April 20). Large Scale Automated Reading of Frontal and Lateral Chest X-Rays using Dual Convolutional Neural Networks. *Machine Intelligence in Medical Imaging*.

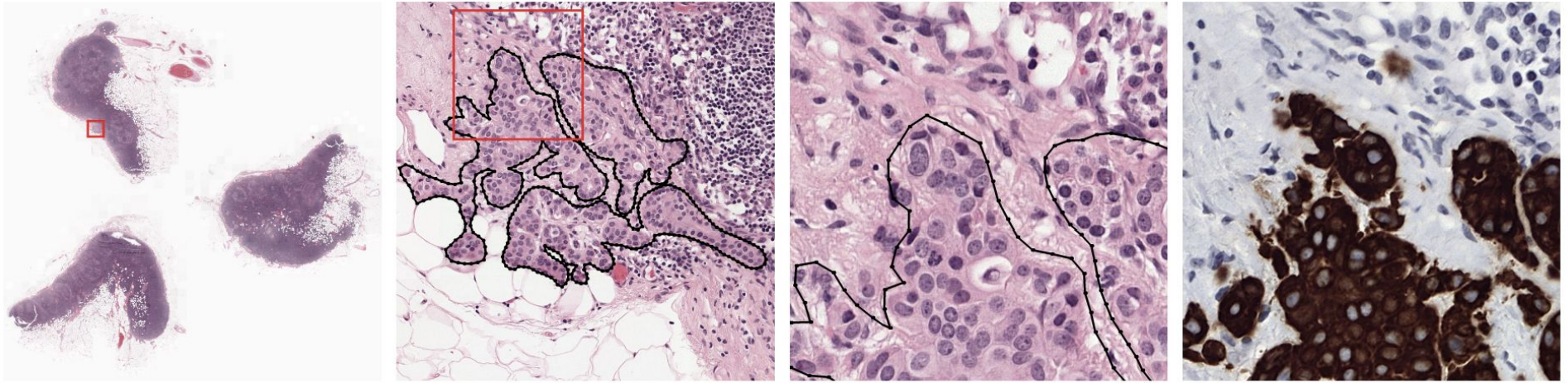
Cohen, J. P., Hashir, M., Brooks, R., & Bertrand, H. (2020). On the limits of cross-domain generalization in automated X-ray prediction. *Medical Imaging with Deep Learning*. https://arxiv.org/abs/2002.02497

Cohen, J. P., Viviano, J., Hashir, M., & Bertrand, H. (2020). TorchXRayVision: A library of chest X-ray datasets and models. *https://github.com/mlmed/torchxrayvision*

Chapter 2

Histology and Segmentation

CAMELYON17: A large high resolution open histology dataset for cancer detection



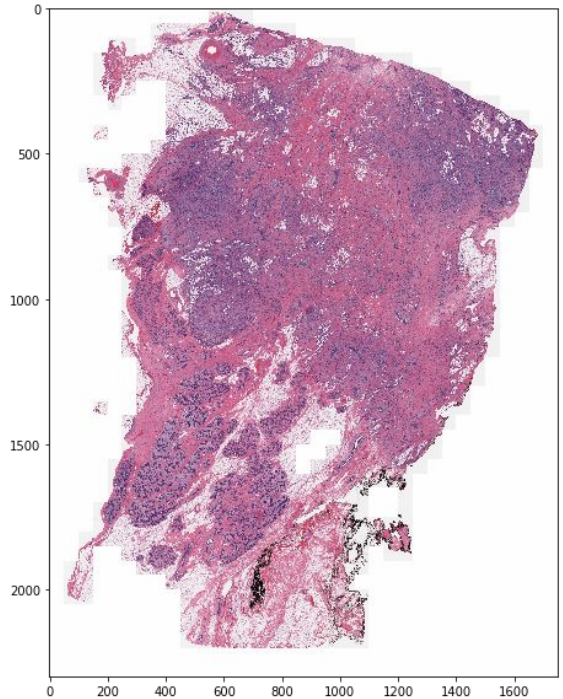
Example of a WSI of a H&E stained section with a delineated micro-metastasis at increasing zoom levels, and the corresponding IHC (cytokeratin 8-18 stained) slide at the same location. The metastasis is outlined with black.

CAMELYON17 Dataset
1000 whole-slide images (WSIs) of sentinel lymph node. (~3GB each!)
5 medical centers. 40 patients from each center. 5 whole-slide images per patient.

Peter Bandi, et al. From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. IEEE-TMI 2018

Patch wise segmentation

Use case: Invasive Ductal Carcinoma (most common subtype of all breast cancers)



Starting with a full slide image of breast tissue.

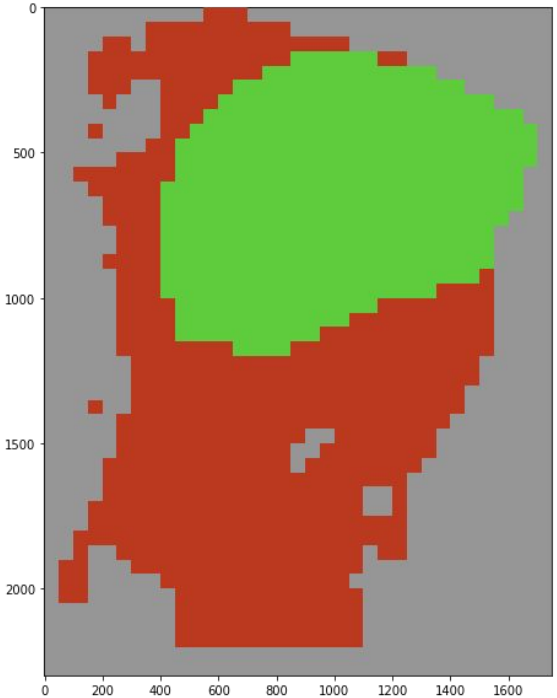


Image is labelled as IDC or not

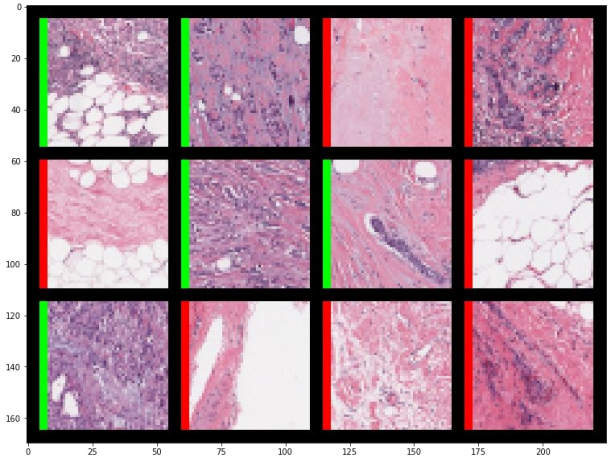
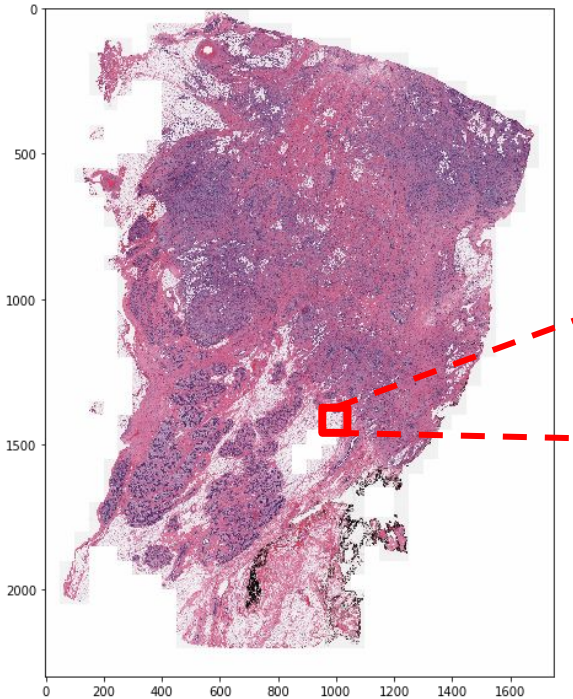


Image is chopped into patches and labelled as IDC or not

https://colab.research.google.com/drive/13T9s3weexAw6YskKoY6c-VvoUgUvWsqf

Patch wise segmentation

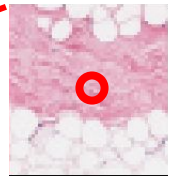
Use case: Invasive Ductal Carcinoma (most common subtype of all breast cancers)



Extract patch

Run through a CNN

Classify center pixel

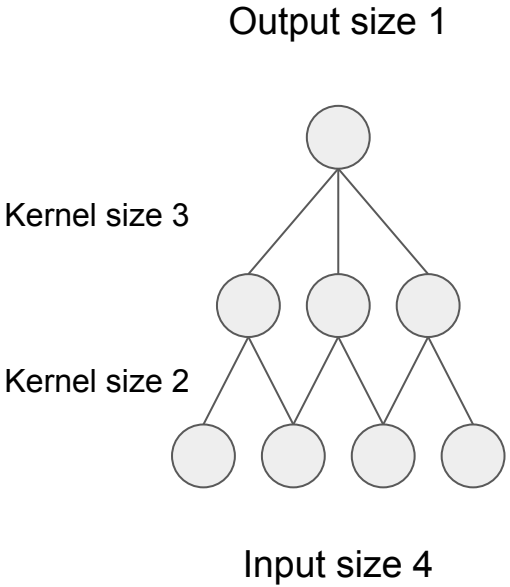


CNN

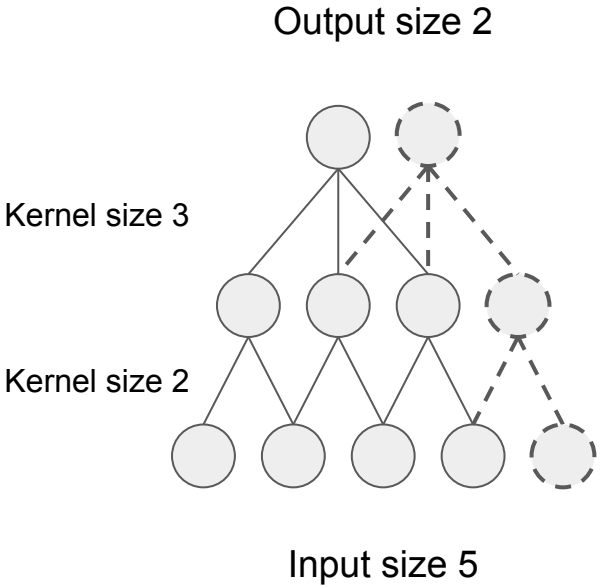
p(cancer)

Class imbalance is an issue. Patch wise training allows easy balancing of classes using standard methods.

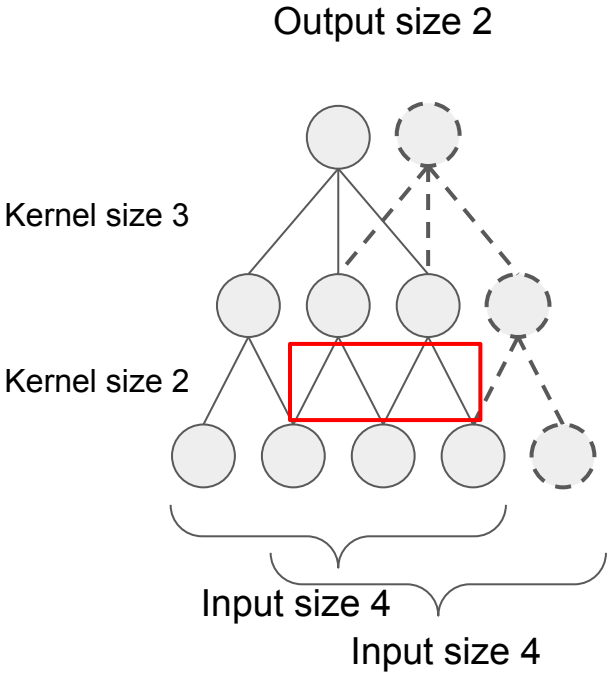
Fully convolutional processing



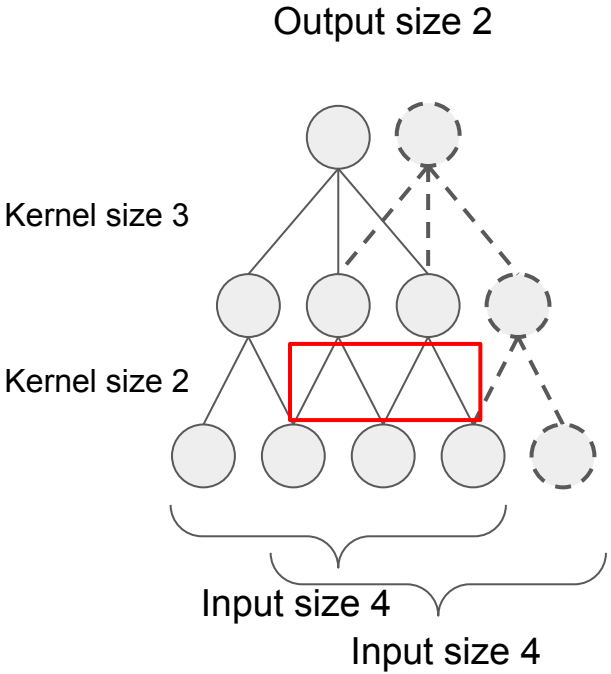
Fully convolutional processing



Fully convolutional processing



Fully convolutional processing



Model's receptive field = 4 nodes

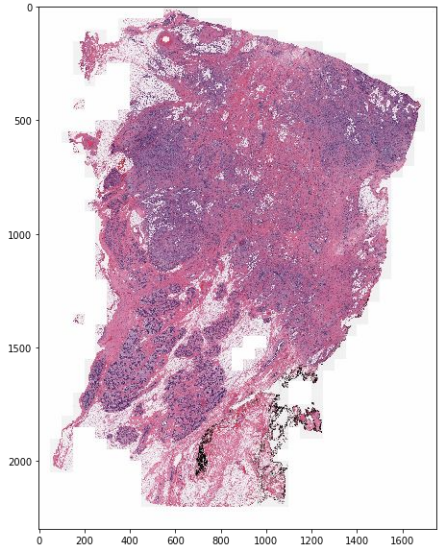
Multiplications saved = 4

Allows for very fast inference.

However, training this way requires a lot of memory. Need to save past outputs.

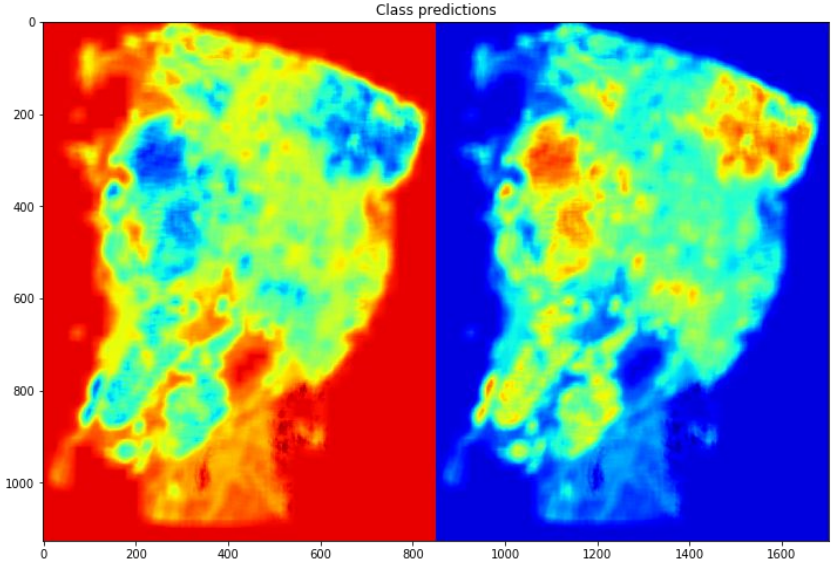
Patch wise training together with FCN inference is a good balance.

Input image

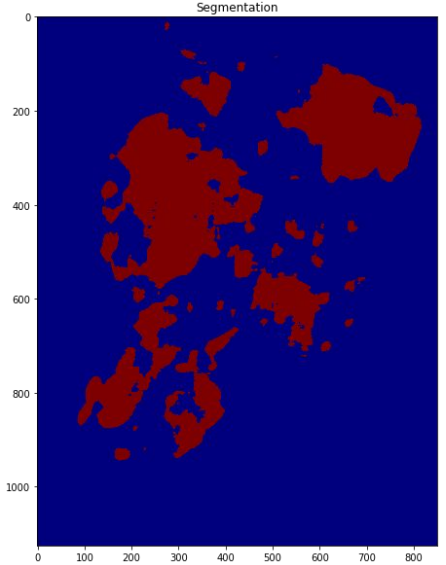


Output class 0

Output class 1

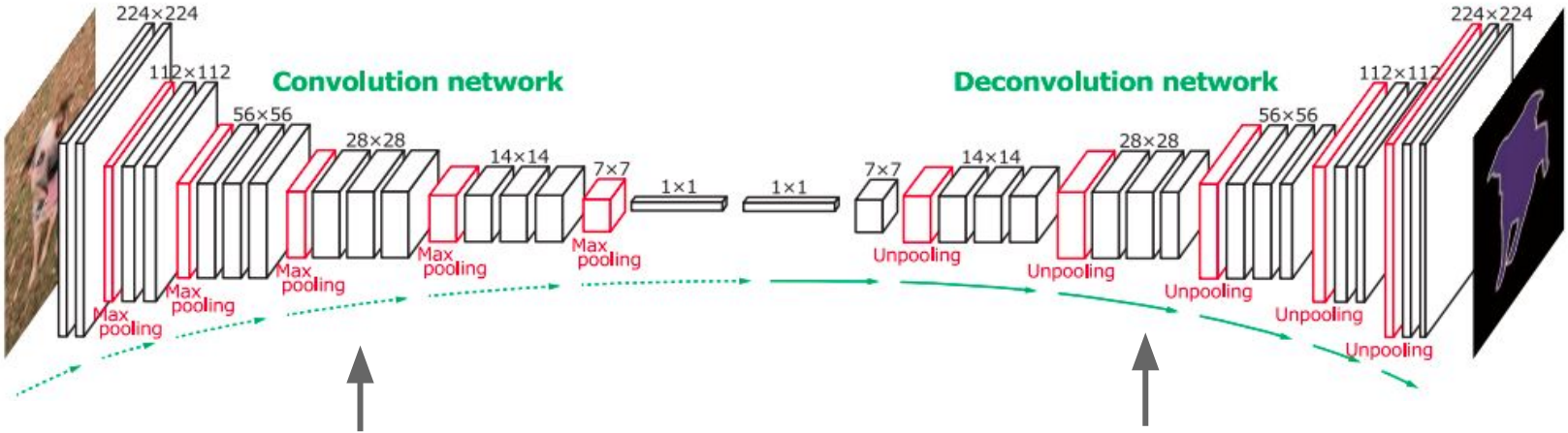


class 1 > class 0



https://colab.research.google.com/drive/13T9s3weexAw6YskKoY6c-VvoUgUvWsqf

Recap: Segmentation using a bottleneck



Normal VGG

“Upside down” VGG

Upsampling possible with

- Unpooling
- Transposed convolutions

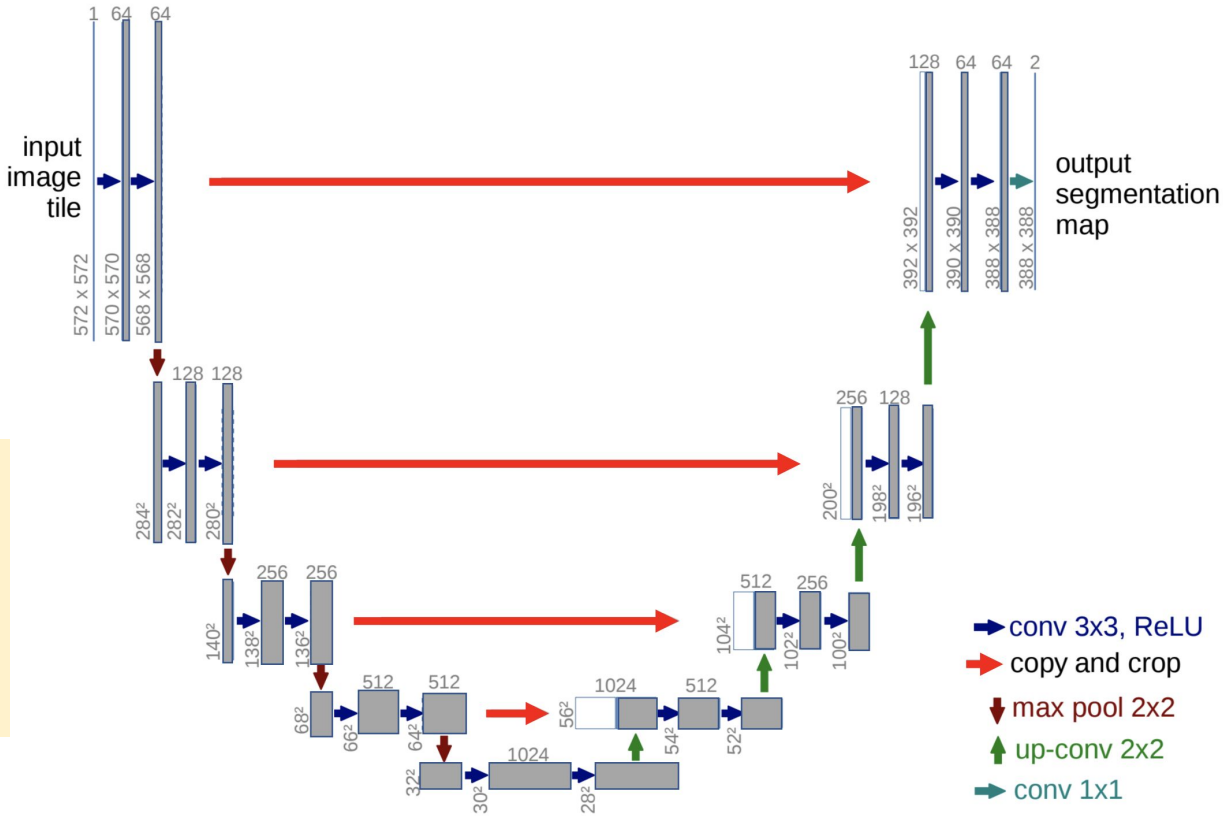
Noh et al, “Learning Deconvolution Network for Semantic Segmentation”, ICCV 2015

Slide design: Fei-Fei Li & Andrej Karpathy & Justin Johnson

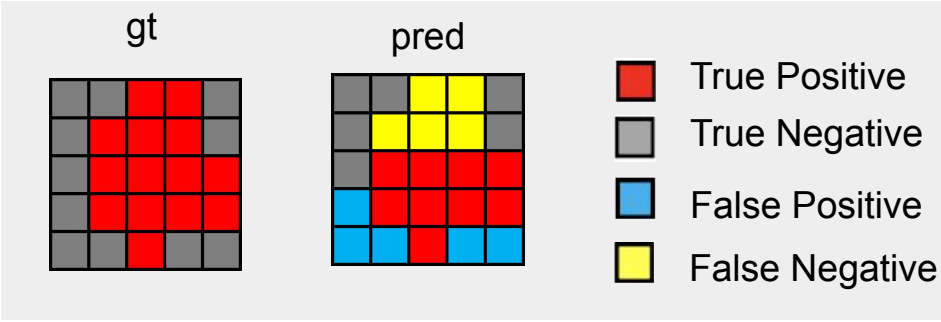
Recap: U-NET

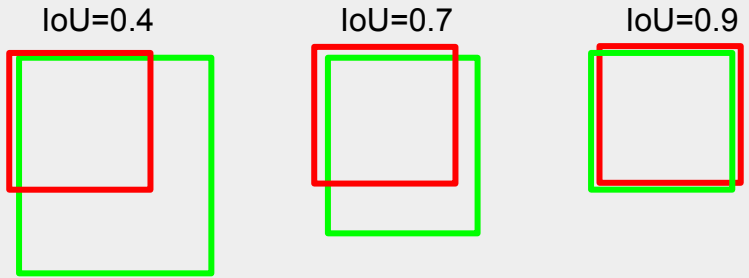
Difference:
Skip connections (like resnet)

Dogma: skips carry spatial information, bottleneck carries high level structure.



Segmentation metrics





$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{IoU} = \text{Jaccard Index} = \frac{TP}{TP + FN + FP} = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

$$\text{Dice Coefficient} = \frac{2 \cdot TP}{(2 \cdot TP) + FN + FP} = \frac{2|X \cap Y|}{|X| + |Y|}$$

Training with dice

Exercise: What p maximizes this?

Using the dot product to compute the intersection allows for a differentiable loss.

$$\mathrm{DL}(p, \hat{p}) = \frac{2 \sum_i p_i \hat{p}_i}{\sum_i (p_i + \hat{p}_i)}$$

For multiple classes a basic approach is to average over all classes

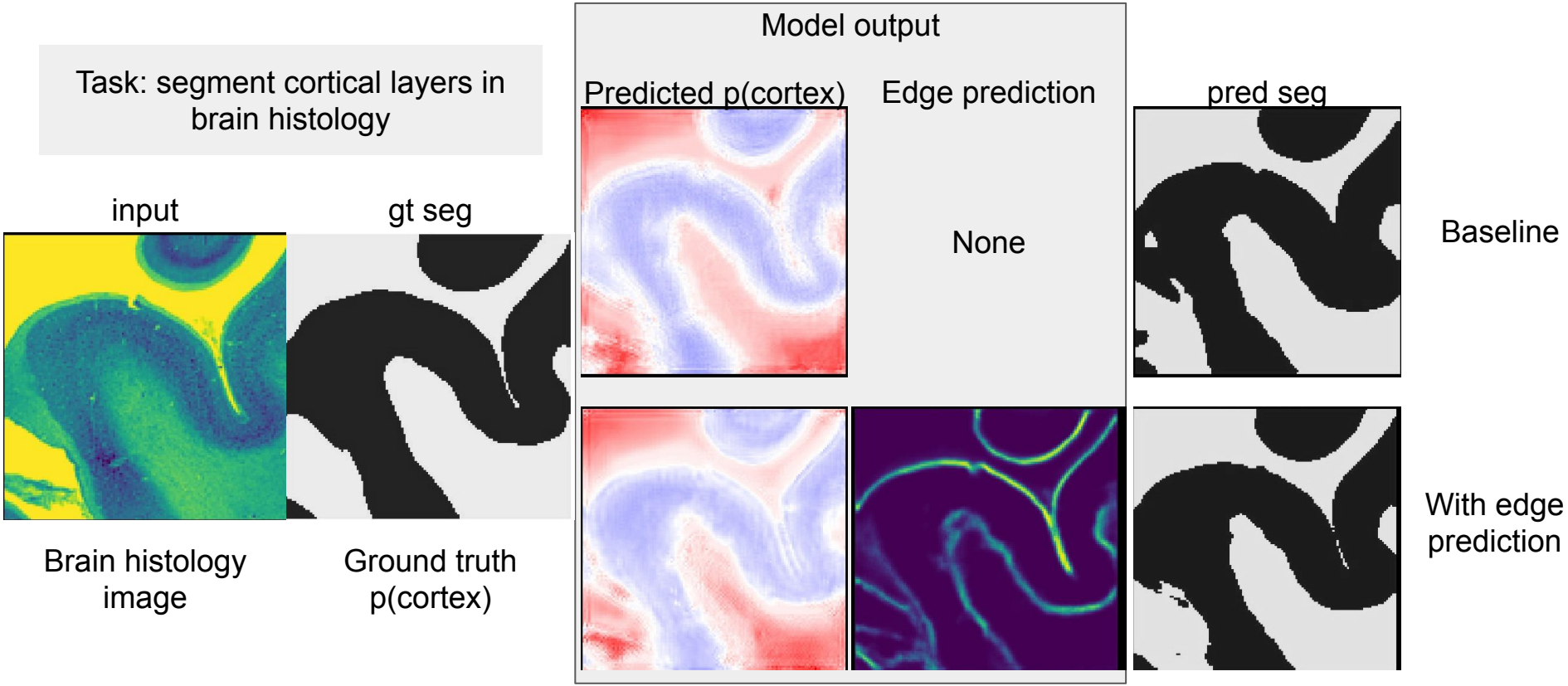
$$\mathrm{DL}_{mean}(p, \hat{p}) = \frac{1}{|C|} \sum_{c \in C} \frac{2 \sum_i p_i^c \hat{p}_i^c}{\sum_i (p_i^c + \hat{p}_i^c)}$$

Use a sigmoid or a softmax to restrict output.

More reading: https://arxiv.org/abs/1707.00478

Tricks: Improving edges in segmentations by predicting edges

Task: segment cortical layers in brain histology



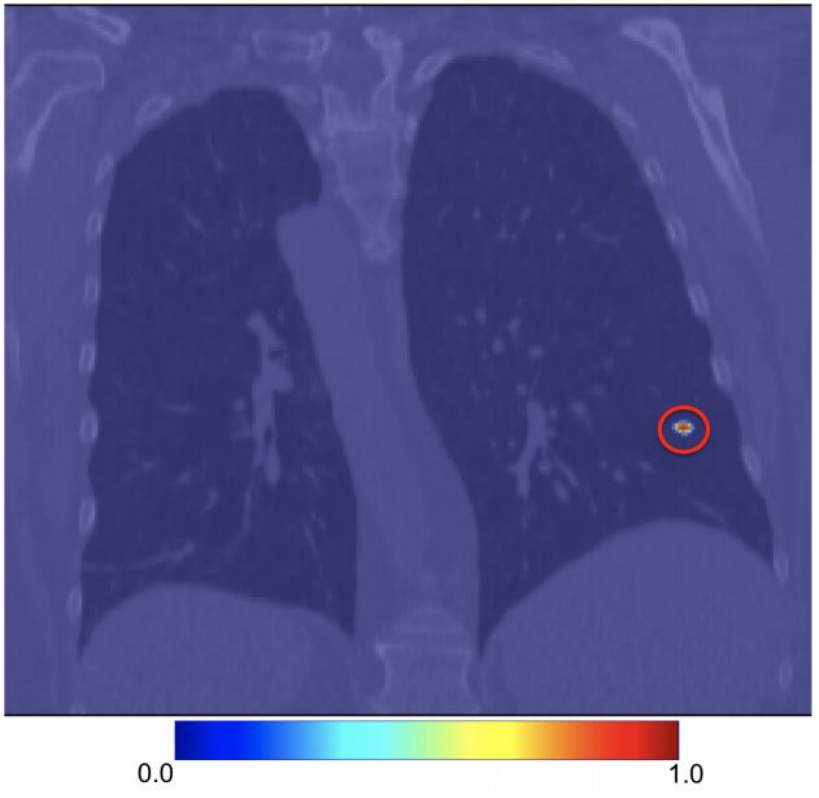
Images provided by Konrad Wagstyl (University College London) 2020

More reading about idea: [Polzounov, WordFence: Text Detection in Natural Images with Border Awareness, 2017]

Challenge: extreme class imbalance (e.g. lung nodule)

Background classes can dominates the loss and cause learning instability do to large gradients.

Balanced sampling may not work as well because patches which could yield false positives are rarely seen to train on.



CASED importance sampling for large images

General Idea:

Store a probability for each patch.

Generate patches based on this probability.

Probability is inverse of how well your model performs on that patch.

Samples are stratified by class.

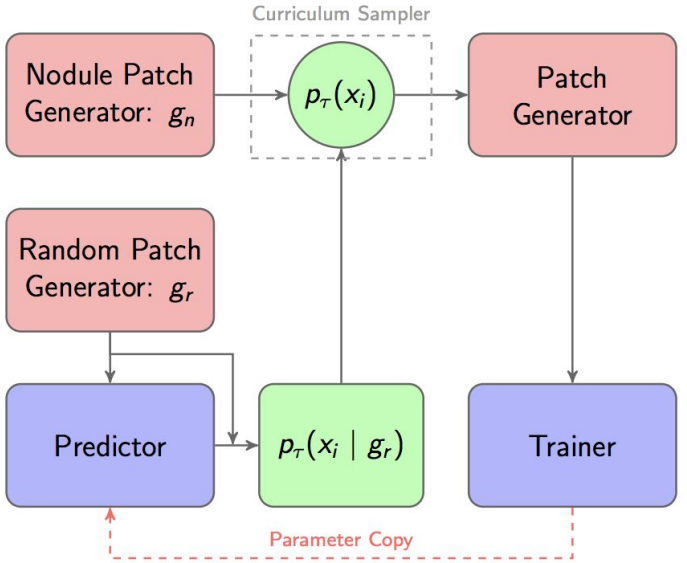


Fig. 1. Schematic diagram of CASED framework

[Jesson, https://arxiv.org/abs/1807.10819]

Chapter 2 - References

Fidon, L., Li, W., Garcia-Peraza-Herrera, L. C., Ekanayake, J., Kitchen, N., Ourselin, S., & Vercauteren, T. (2017). *Generalised Wasserstein Dice Score for Imbalanced Multi-class Segmentation using Holistic Convolutional Networks*. http://arxiv.org/abs/1707.00478

Litjens, G., Bandi, P., Bejnordi, B. E., et al (2018). 1399 H&E-stained sentinel lymph node sections of breast cancer patients: The CAMELYON dataset. In *GigaScience* (Vol. 7, Issue 6). https://doi.org/10.1093/gigascience/giy065

Noh, H., Hong, S., & Han, B. (2015). Learning Deconvolution Network for Semantic Segmentation. *International Conference on Computer Vision*. https://arxiv.org/abs/1505.04366

Polzounov, A., Ablavatski, A., Escalera, S., Lu, S., & Cai, J. (2018). Wordfence: Text detection in natural images with border awareness. *International Conference on Image Processing*, *2017-Septe*, 1222–1226. https://doi.org/10.1109/ICIP.2017.8296476

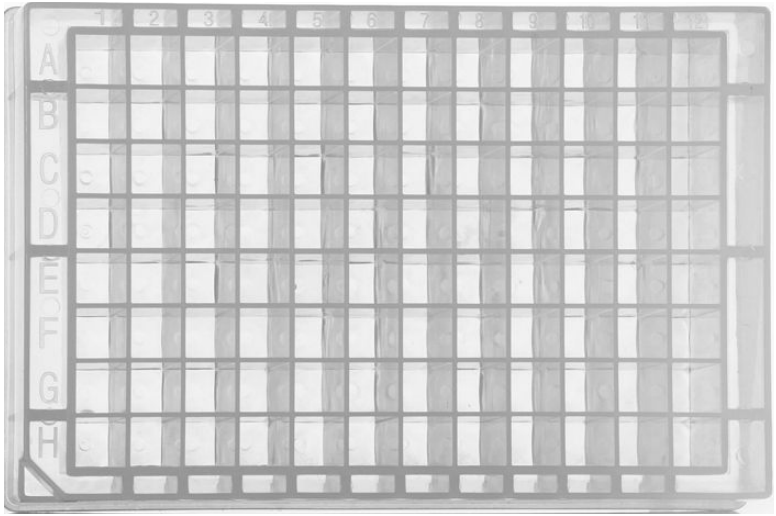
Jesson, A., Guizard, N., Ghalehjegh, S. H., Goblot, D., Soudan, F., & Chapados, N. (2017, September 10). CASED: Curriculum Adaptive Sampling for Extreme Data Imbalance. *Medical Image Computing and Computer Assisted Intervention*. https://doi.org/10.1007/978-3-319-66179-7_73

Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of Pathology Informatics, 7(1), 29. https://doi.org/10.4103/2153-3539.186902

Chapter 3

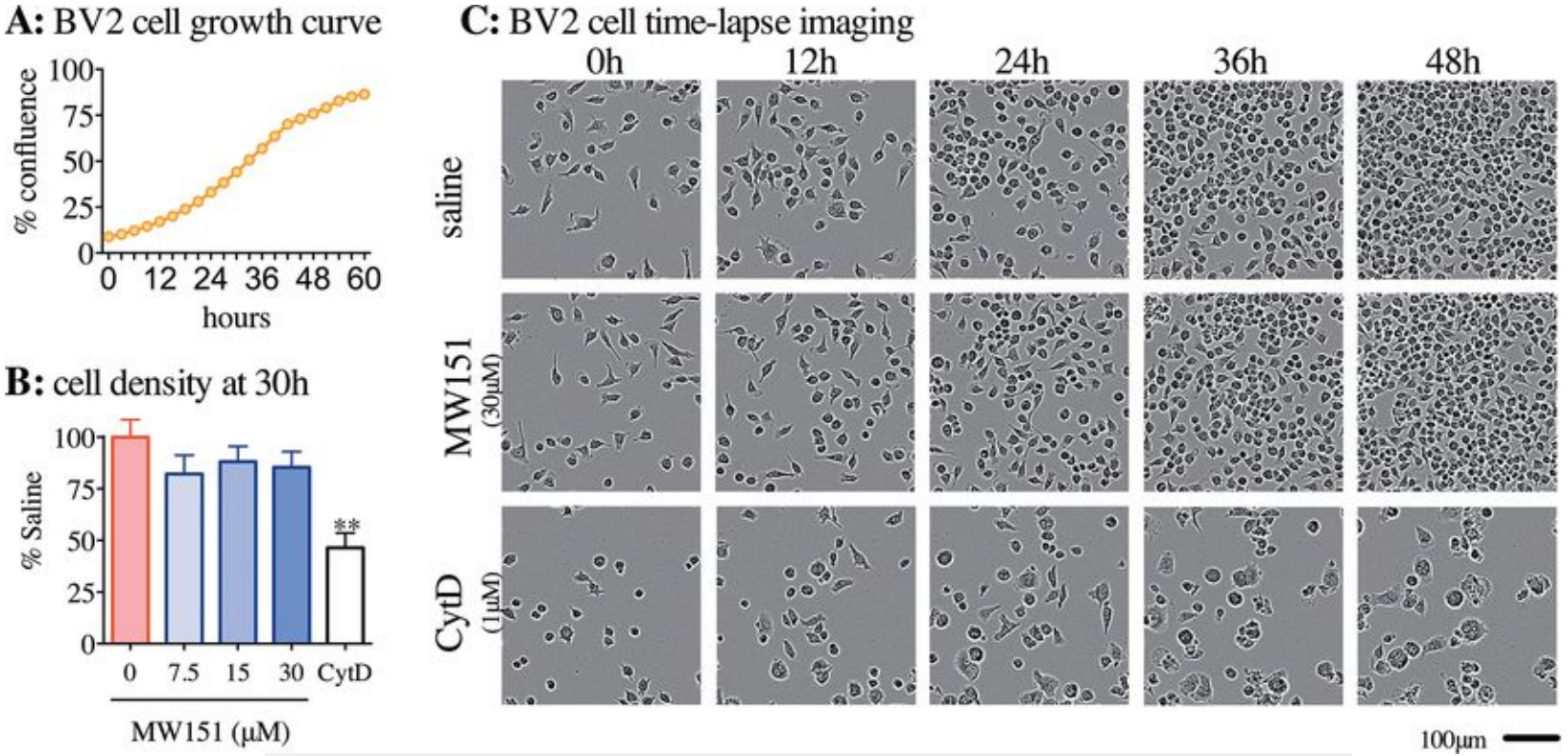
Cell Counting

Use case: Proliferation/Cell growth studies



Standard 96-well plate

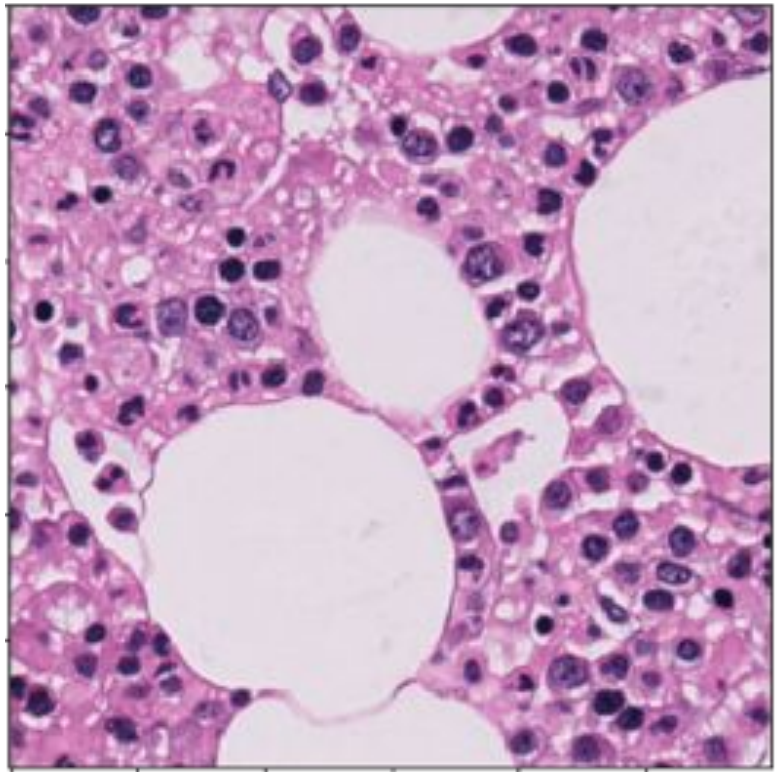
Treat cells with different compounds and observe proliferation over time

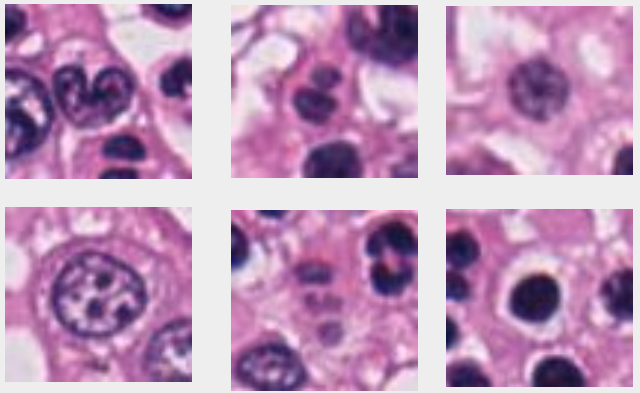


Bachstetter, MW151 Inhibited IL-1? Levels after Traumatic Brain Injury with No Effect on Microglia Physiological Responses, PLOS ONE, 2017

Use case: Proliferation/Cell growth studies

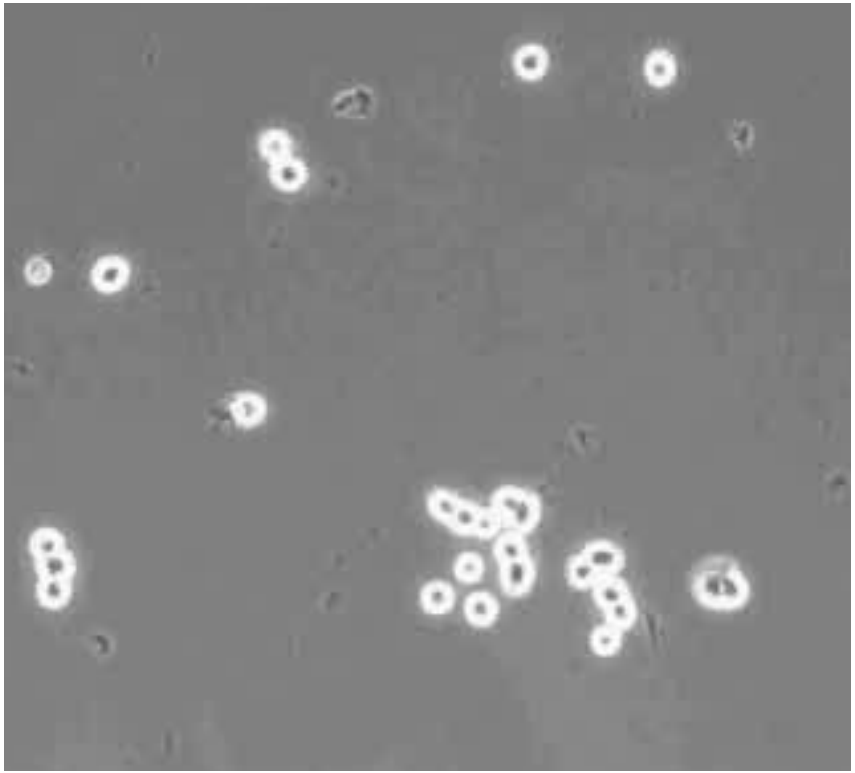
Use case: Counting in histology slides





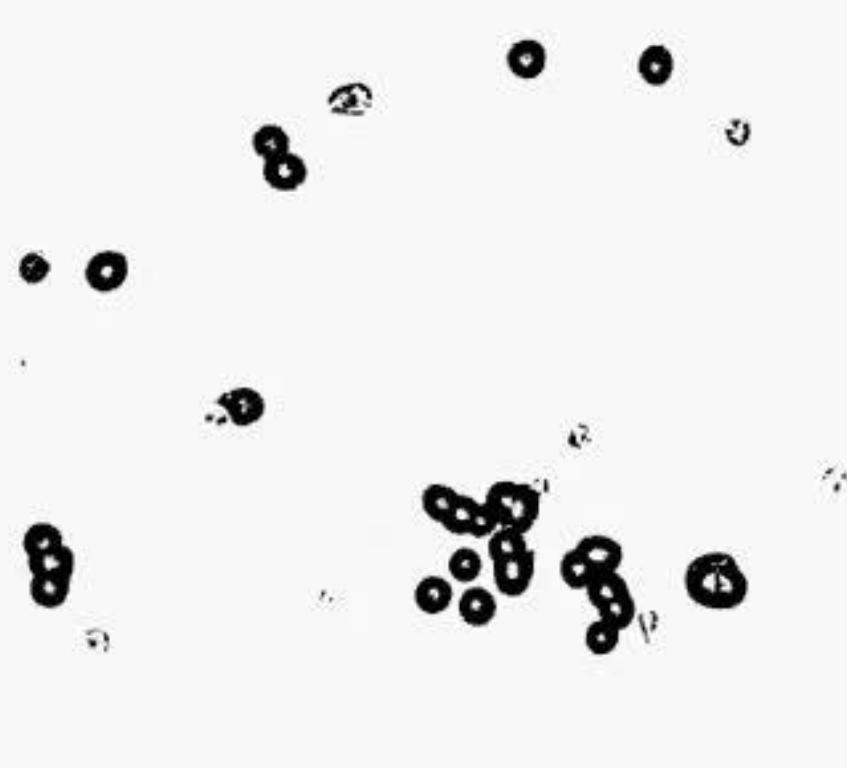
Complicated cell structure

Cell counting (classic CV)



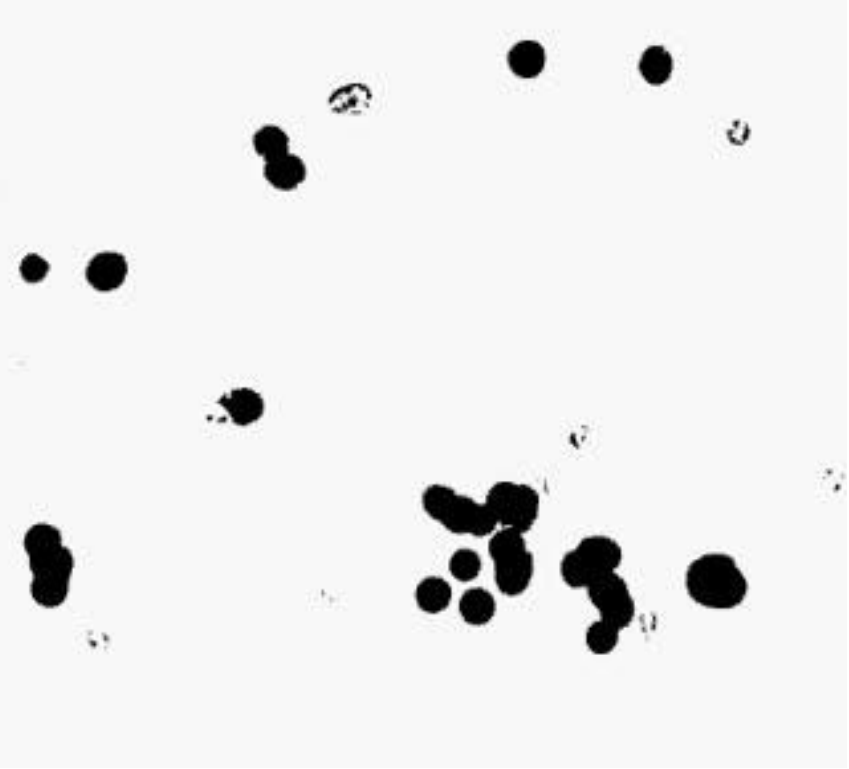
1. Create binary segmentation image
2. Watershed segmentation
3. Isolate and count

Cell counting (classic CV)



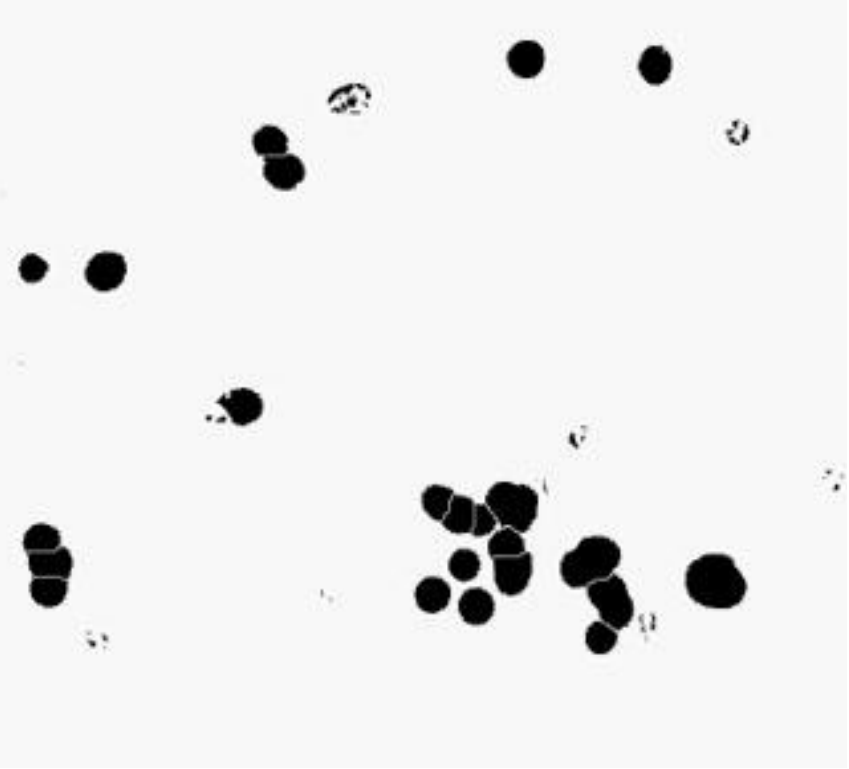
1. **Create binary segmentation image**
2. Watershed segmentation
3. Isolate and count

Cell counting (classic CV)



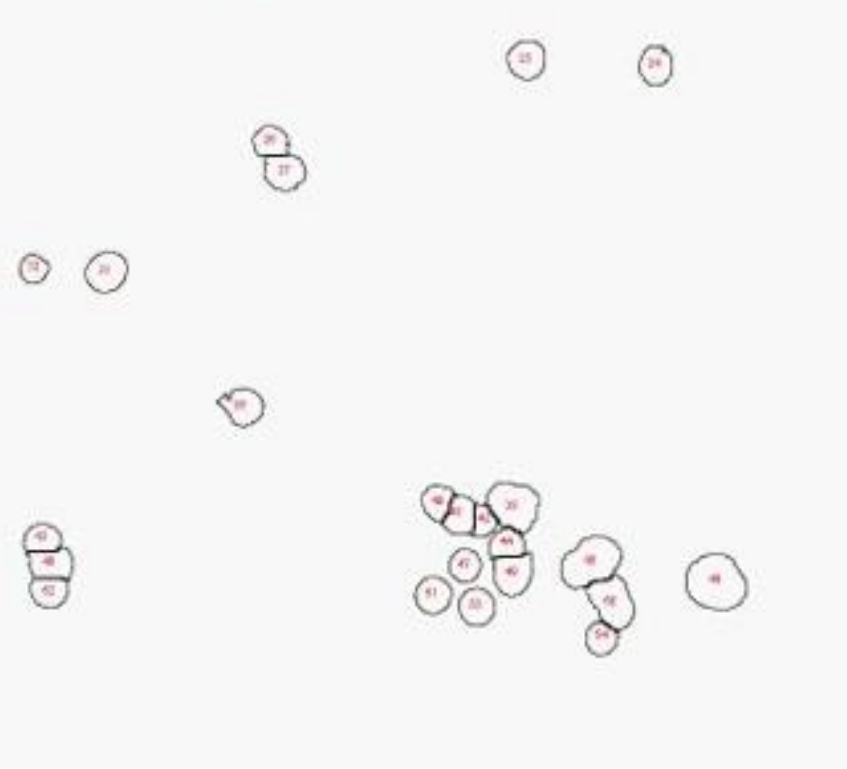
1. **Create binary segmentation image**
2. Watershed segmentation
3. Isolate and count

Cell counting (classic CV)



1. Create binary segmentation image
2. **Watershed segmentation**
3. Isolate and count

Cell counting (classic CV)

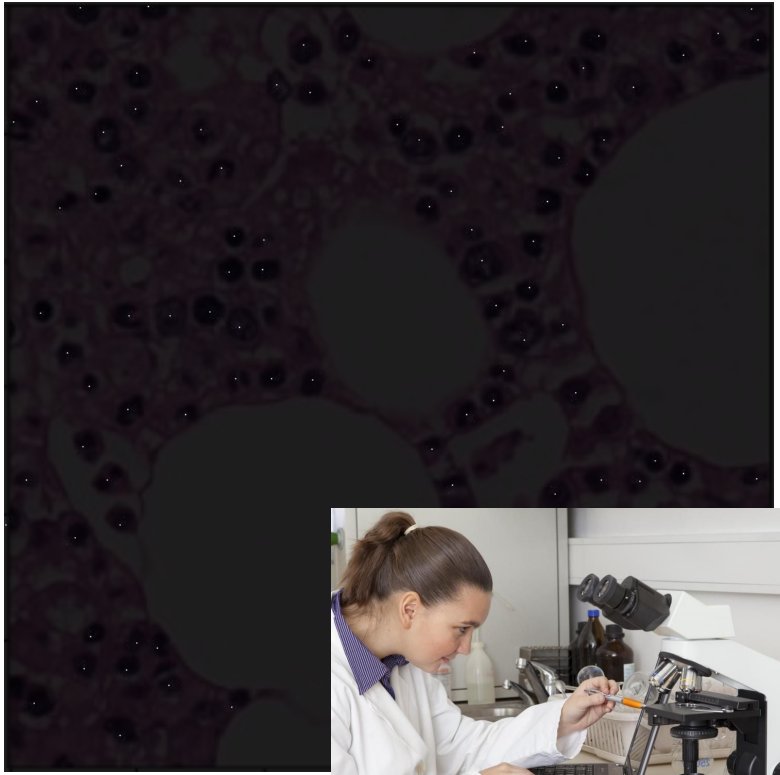


1. Create binary segmentation image
2. Watershed segmentation
3. **Isolate and count**

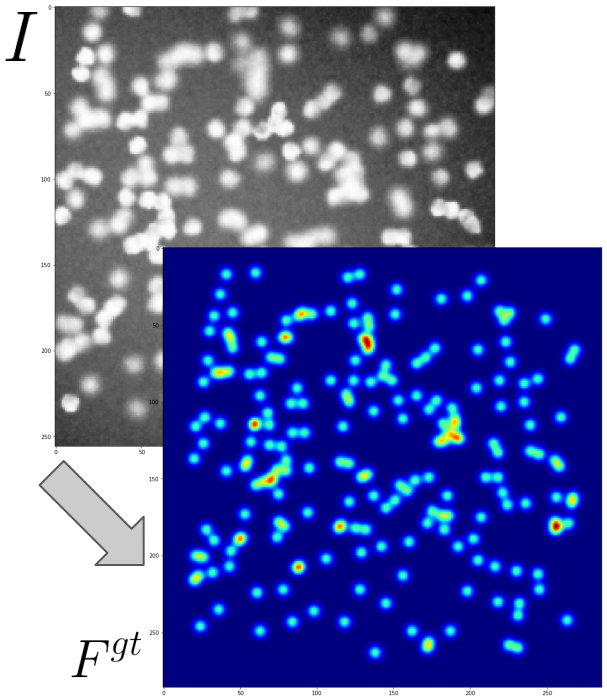
This works well on easy tasks but doesn't scale.

"Pipelines" end up breaking on new images with different lighting or stain.

How to get labels?



Counting via Segmentation



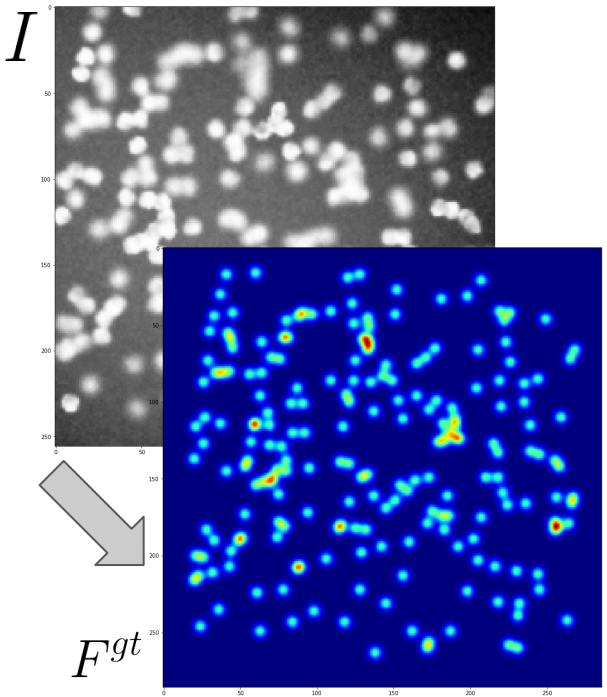
Targets for regression

$$F^{gt}(x,y) = \sum_{i}^{\#\text{ Cells}} \mathcal{N}([x,y];[x_i,y_i],\sigma^2)$$

Sigma is typically small like a few pixels

V. Lempitsky and A. Zisserman, “Learning To Count Objects in Images,” **2010**.

Counting via Segmentation



Targets for regression

$$F^{gt}(x,y) = \sum_{i}^{\#\text{ Cells}} \mathcal{N}([x,y];[x_i,y_i],\sigma^2)$$

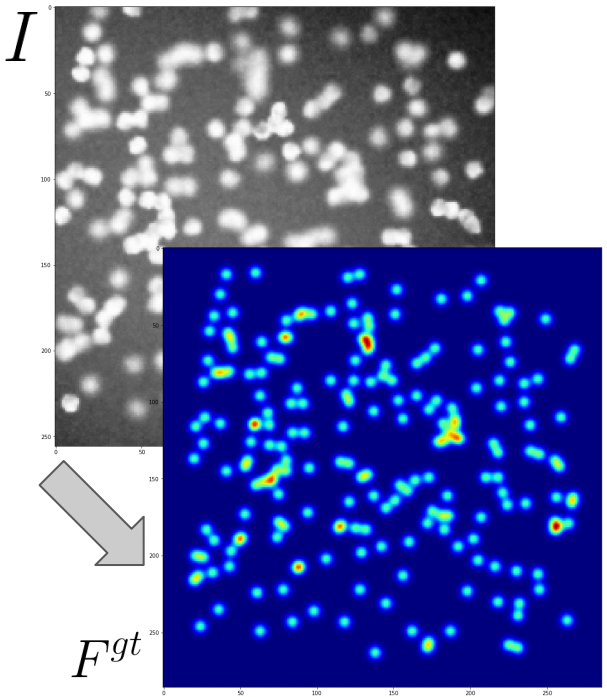
Sigma is typically small like a few pixels

Train model to regress

$$L = \sum_{x,y} |F^{gt}(x,y) - F(I)(x,y)|$$

V. Lempitsky and A. Zisserman, “Learning To Count Objects in Images,” **2010**.

Counting via Segmentation



Targets for regression

$$F^{gt}(x,y) = \sum_{i}^{\#\text{ Cells}} \mathcal{N}([x,y];[x_i,y_i],\sigma^2)$$

Sigma is typically small like a few pixels

Train model to regress

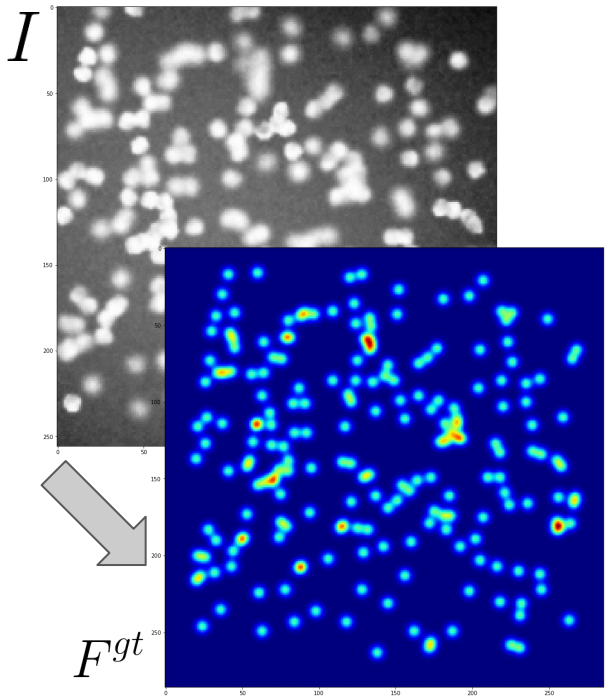
$$L = \sum_{x,y} |F^{gt}(x,y) - F(I)(x,y)|$$

To recover count:

$$\text{count} = \sum_{x,y} F(I)(x,y)$$

V. Lempitsky and A. Zisserman, “Learning To Count Objects in Images,” **2010**.

Counting via Segmentation



Targets for regression

$$F^{gt}(x,y) = \sum_{i}^{\#\text{ Cells}} \mathcal{N}([x,y];[x_i,y_i],\sigma^2)$$

Sigma is typically small like a few pixels

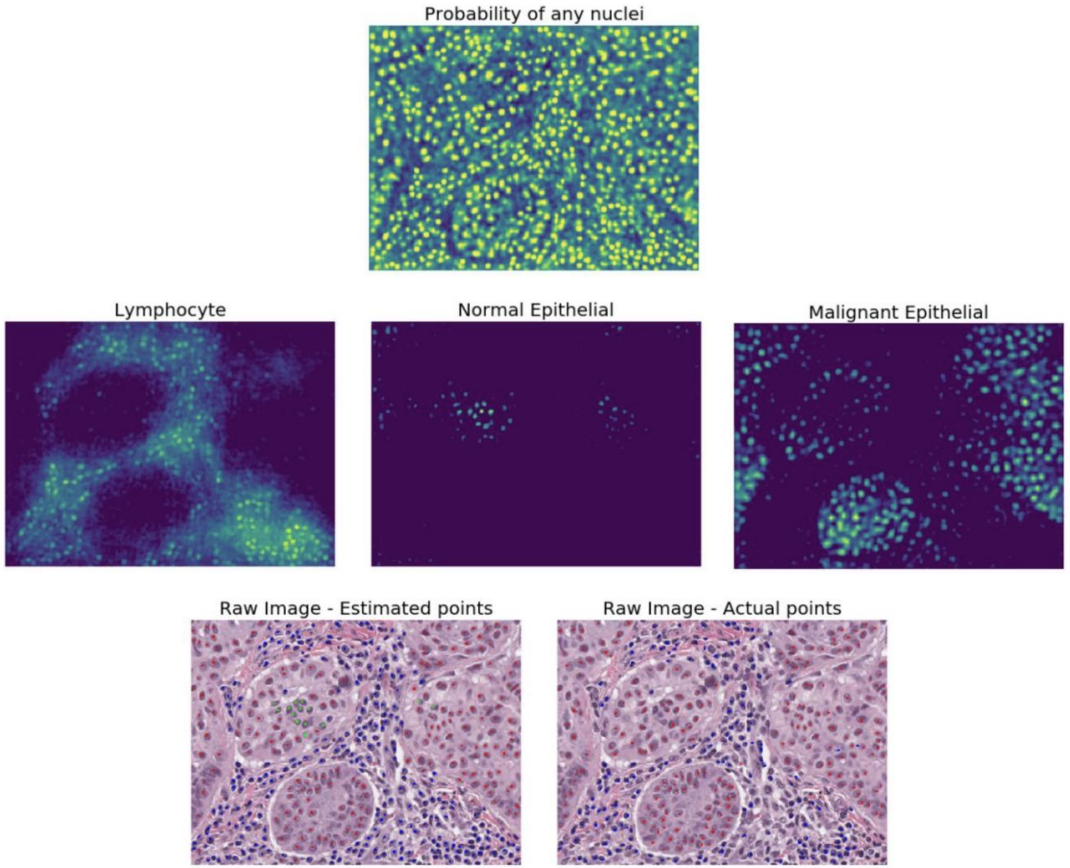
Note: Square kernels for redundant counting work better [Cohen 2017]

To recover count:

$$\text{count} = \sum_{x,y} F(I)(x,y)$$

V. Lempitsky and A. Zisserman, “Learning To Count Objects in Images,” **2010**.

Multiple output classes



Count and classify different cell types [Bidart 2018]

Counting and classifying also possible using multiple output channels.

Combine losses together

$$\lambda_1 L_{lymph} + \lambda_2 L_{norm} + \lambda_3 L_{mal}$$

Max prediction over output channels for each cell identified

Chapter 3 - References

Lempitsky, V., & Zisserman, A. (2010). Learning To Count Objects in Images. *Neural Information Processing Systems (NeurIPS)*.

Cohen, J. P., Boucher, G., Glastonbury, C. A., Lo, H. Z., & Bengio, Y. (2017). Count-ception: Counting by Fully Convolutional Redundant Counting. *International Conference on Computer Vision Workshop on BioImage Computing*. http://arxiv.org/abs/1703.08710

Xie, W., Noble, J. A., & Zisserman, A. (2016). Microscopy cell counting and detection with fully convolutional regression networks. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. https://doi.org/10.1080/21681163.2016.1149104

Gangeh, M. J., Bidart, R., Peikari, M., Martel, A. L., Ghodsi, A., Salama, S., & Nofech-Mozes, S. (2018). Localization and classification of cell nuclei in post-neoadjuvant breast cancer surgical specimen using fully convolutional networks. In M. N. Gurcan & J. E. Tomaszewski (Eds.), *Medical Imaging 2018: Digital Pathology* (Vol. 10581, p. 23). SPIE. https://doi.org/10.1117/12.2292815

BBBC021 - Human MCF7 cells – compound-profiling

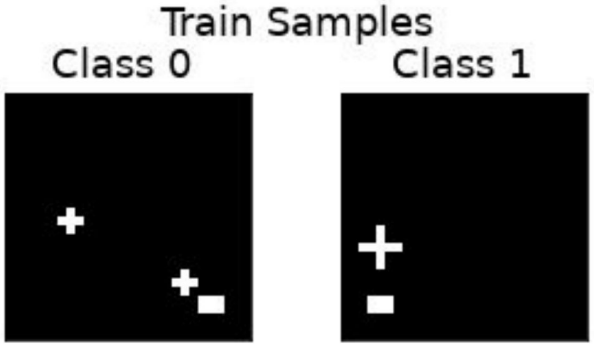
RxRx1 - CellSignal: Disentangling biological signal from experimental noise

MBM - Modified Bone Marrow cell counting dataset

Chapter 4

Incorrect Feature Attribution

Incorrect feature attribution



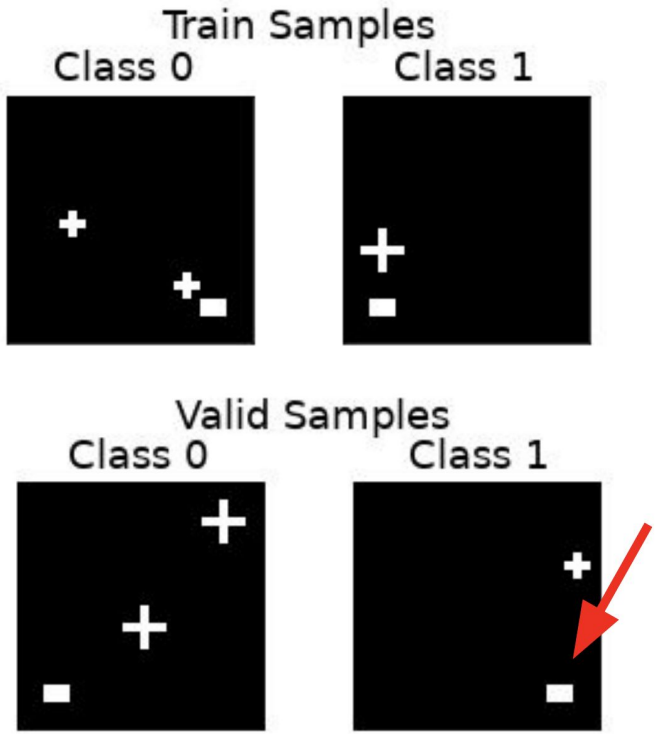
Goal: predict if there are two plus signs anywhere

However, an easy to spot confounder exists!

The confounding variable distracts the model causing it to fail to generalize.

[Ross, Right for the Right Reasons, 2017]
[Viviano, Underwhelming Generalization Improvements From Controlling Feature Attribution, 2019]

Incorrect feature attribution



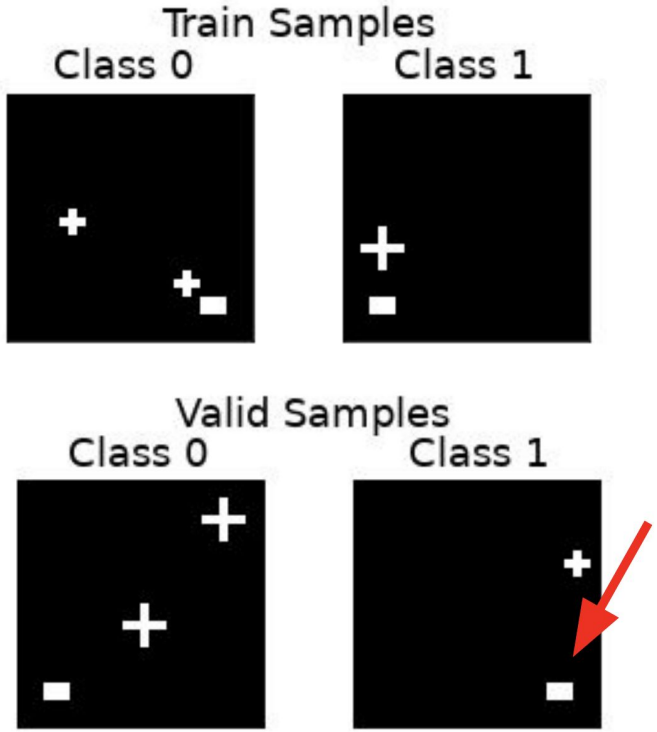
Goal: predict if there are two plus signs anywhere

However, an easy to spot confounder exists!

The confounding variable distracts the model causing it to fail to generalize.

[Ross, Right for the Right Reasons, 2017]
[Viviano, Underwhelming Generalization Improvements From Controlling Feature Attribution, 2019]

Incorrect feature attribution



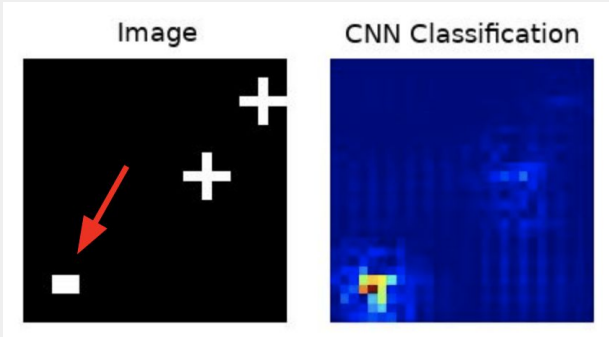
Goal: predict if there are two plus signs anywhere

However, an easy to spot confounder exists!

The confounding variable distracts the model causing it to fail to generalize.

We can observe this by looking at the saliency map

$$\left|\frac{\partial f}{\partial x}\right|$$



[Ross, Right for the Right Reasons, 2017]
[Viviano, Underwhelming Generalization Improvements From Controlling Feature Attribution, 2019]

Incorrect feature attribution

Models can overfit to confounding variables in the data.

- Merging datasets with different class imbalance (confounding artifacts from each hospital)
- Labels confounding with each other
- Demographics confounding with labels

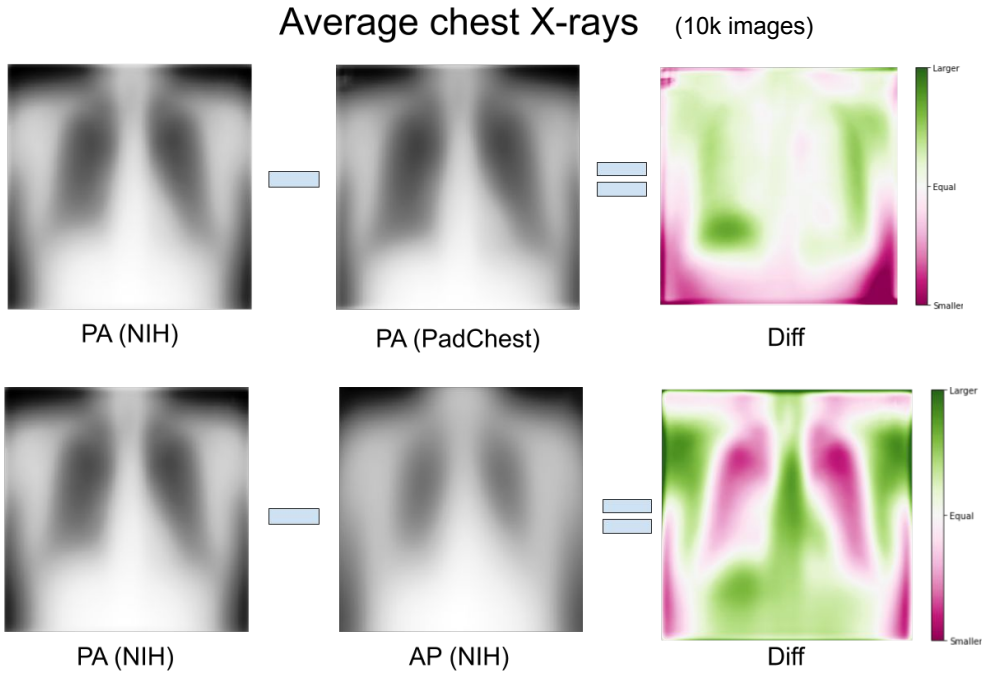
[Ross, Right for the Right Reasons, 2017]
[Zeck, Confounding variables can degrade generalization performance of radiological ..., 2018]
[Viviano, Underwhelming Generalization Improvements From Controlling Feature Attribution, 2019]
[Simpson, GradMask: Reduce Overfitting by Regularizing Saliency, 2019]

Incorrect feature attribution

Models can overfit to confounding variables in the data.

- Merging datasets with different class imbalance (confounding artifacts from each hospital)
- Labels confounding with each other
- Demographics confounding with labels

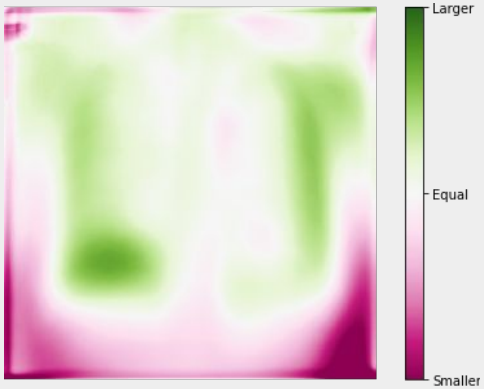
Example:Systematic discrepancy between average image in datasets



[Ross, Right for the Right Reasons, 2017]
[Zeck, Confounding variables can degrade generalization performance of radiological ..., 2018]
[Viviano, Underwhelming Generalization Improvements From Controlling Feature Attribution, 2019]
[Simpson, GradMask: Reduce Overfitting by Regularizing Saliency, 2019]

Incorrect feature attribution

Recall:
NIH/PADCHEST Diff



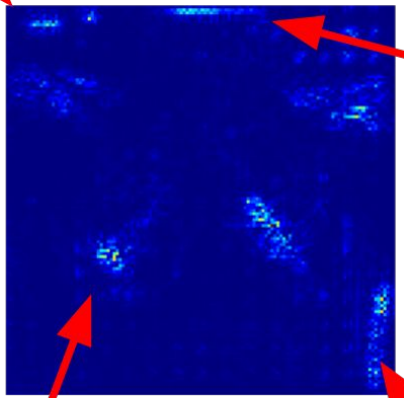
Symptoms of incorrect feature attribution

$$\left|\frac{\partial f}{\partial x}\right|$$

Image text should not change prediction

Input Image

Gradient Saliency Map



Neck edge has high contrast and could be a proxy for a subset of patients

Lung visibility seen as difference between average images

Corner is predictive in average images

[Viviano, Underwhelming Generalization Improvements From Controlling Feature Attribution, 2019]

Mitigation approaches

Feature engineering

- **Range normalization** (/max)
- **Subspace alignment** (align data using their eigenbasis based on a feature) [Fernando 2014]
- **Removing the largest principle component** (joint PCA and reconstruct without largest eigenvector)

Mitigation approaches

Feature engineering

- **Range normalization** (/max)
- **Subspace alignment** (align data using their eigenbasis based on a feature) [Fernando 2014]
- **Removing the largest principle component** (joint PCA and reconstruct without largest eigenvector)

During training

- **Reverse gradient** (make intermediate layer invariant to a label) [Ganin & Lempitsky, 2014]
- **Right for the Right Reasons** (regularize saliency map) [Ross, Hughes, & Finale Doshi-Velez, 2017]
- **GradMask** (regularize contrast saliency map between classes) [Simpson, 2019]
- **ActivDiff** (regularize representation to focus on pathology) [Viviano, 2019]

What if feature artifact is correlated with target label?
Is the reason that should be used for prediction known?
What if it is not known?

Chapter 4 - References

Ganin, Y., & Lempitsky, V. (2015, September 26). Unsupervised Domain Adaptation by Backpropagation. Proceedings of the International Conference on Machine Learning (ICML). http://jmlr.org/proceedings/papers/v37/ganin15.html

Ross, A., Hughes, M. C., & Doshi-Velez, F. (2017). Right for the Right Reasons: Training Differentiable Models by Constraining their Explanations. *International Joint Conference on Artificial Intelligence*. https://github.com/dtak/rrr.

Viviano, J. D., Simpson, B., Dutil, F., Bengio, Y., & Cohen, J. P. (2019). Underwhelming Generalization Improvements From Controlling Feature Attribution. *Arxiv:1910.00199*. http://arxiv.org/abs/1910.00199

Simpson, B., Dutil, F., Bengio, Y., & Cohen, J. P. (2019, April 16). GradMask: Reduce Overfitting by Regularizing Saliency. *Medical Imaging with Deep Learning Workshop*. http://arxiv.org/abs/1904.07478

Fernando, B., Habrard, A., Sebban, M., & Tuytelaars, T. (2014). *Subspace Alignment For Domain Adaptation*. http://arxiv.org/abs/1409.5241

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Medicine*, *15*(11). https://doi.org/10.1371/journal.pmed.1002683

Seyyed-Kalantari, L., Liu, G., McDermott, M., & Ghassemi, M. (2020). *CheXclusion: Fairness gaps in deep chest X-ray classifiers*. http://arxiv.org/abs/2003.00827

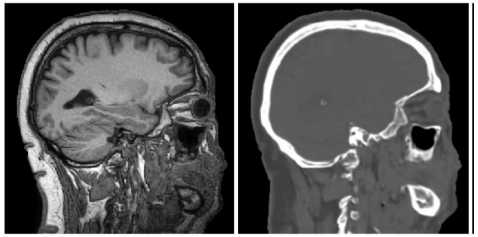
Chapter 5

GANs in Medical Imaging

Medical image-to-image translation considered harmful

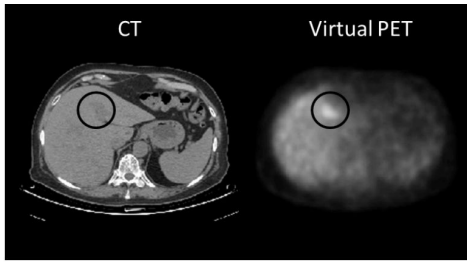
Many papers have proposed methods that can "translate between modalities"

MR -> CT

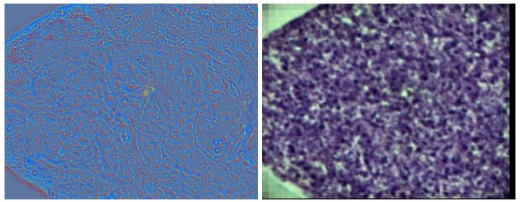


I_{MR} $Syn_{CT}(I_{MR})$

CT -> PET



Synthesized H&E staining



Adversarial losses are very good at distribution matching (e.g. CycleGAN).
But artifacts could be introduced and then used in diagnosis which can be dangerous.

Use case: MRI modality transformation

Undersampled raw MRI source data

Image Translation/ Synthesis

T1 Transformed

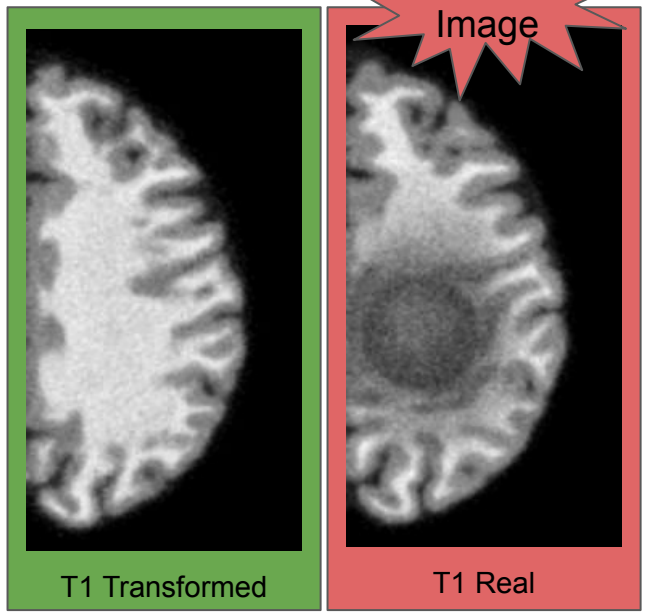
But a bias in training data can lead to incorrect translation

Cohen, Distribution Matching Losses Can Hallucinate Features in Medical Image Translation, 2018

Use case: MRI modality transformation

Undersampled raw MRI source data

Image Translation/ Synthesis

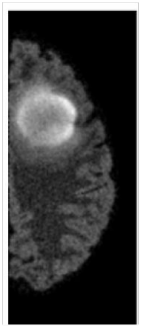


Real Image

Everyone is so healthy!

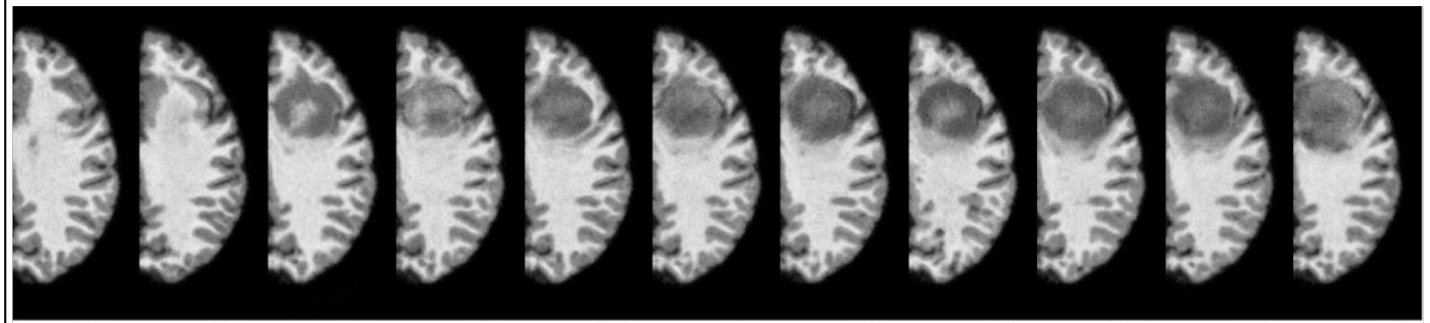
But a bias in training data can lead to incorrect translation

Cohen, Distribution Matching Losses Can Hallucinate Features in Medical Image Translation, 2018

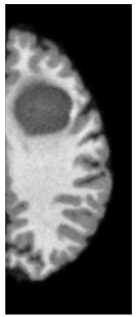


Input
Flair Image

Input image transformed using biased training data



Percentage of training data containing tumors



Ground Truth
T1 Image

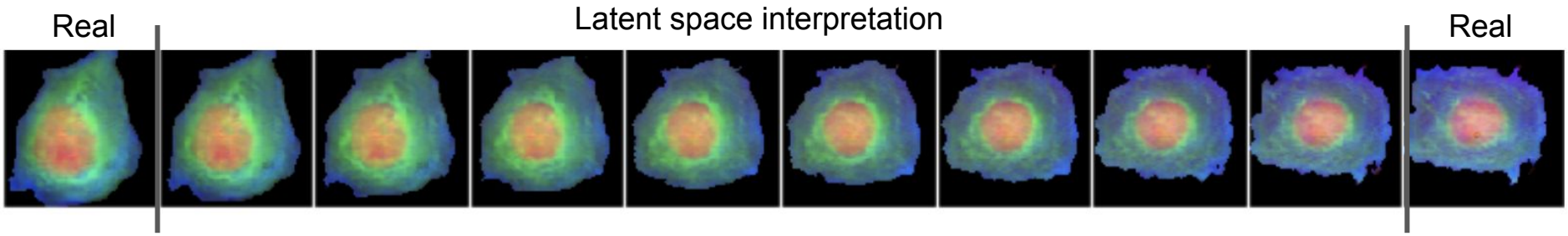
Tumors here are a proxy to illustrate the impact of an unaccounted pathology

Cohen, Distribution Matching Losses Can Hallucinate Features in Medical Image Translation, 2018

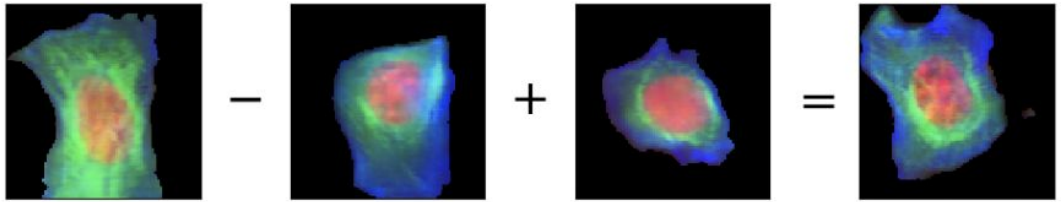
Adversarial losses are useful for representation learning

Example: CytoGAN learning a self-supervised representation for cell images.

- Encoder can be useful for semi-supervised learning
- Exploring representations to understand the cell biology

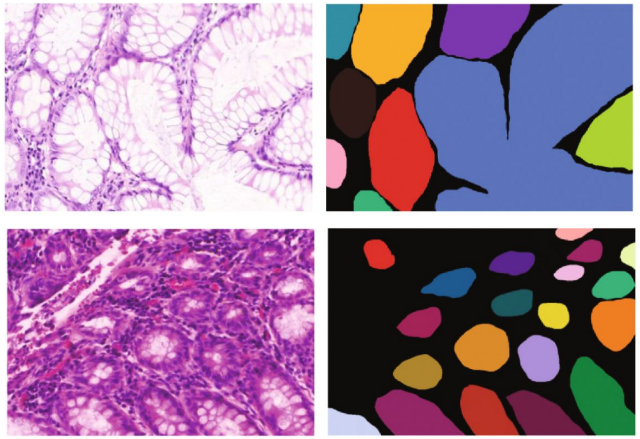


Vector algebra:

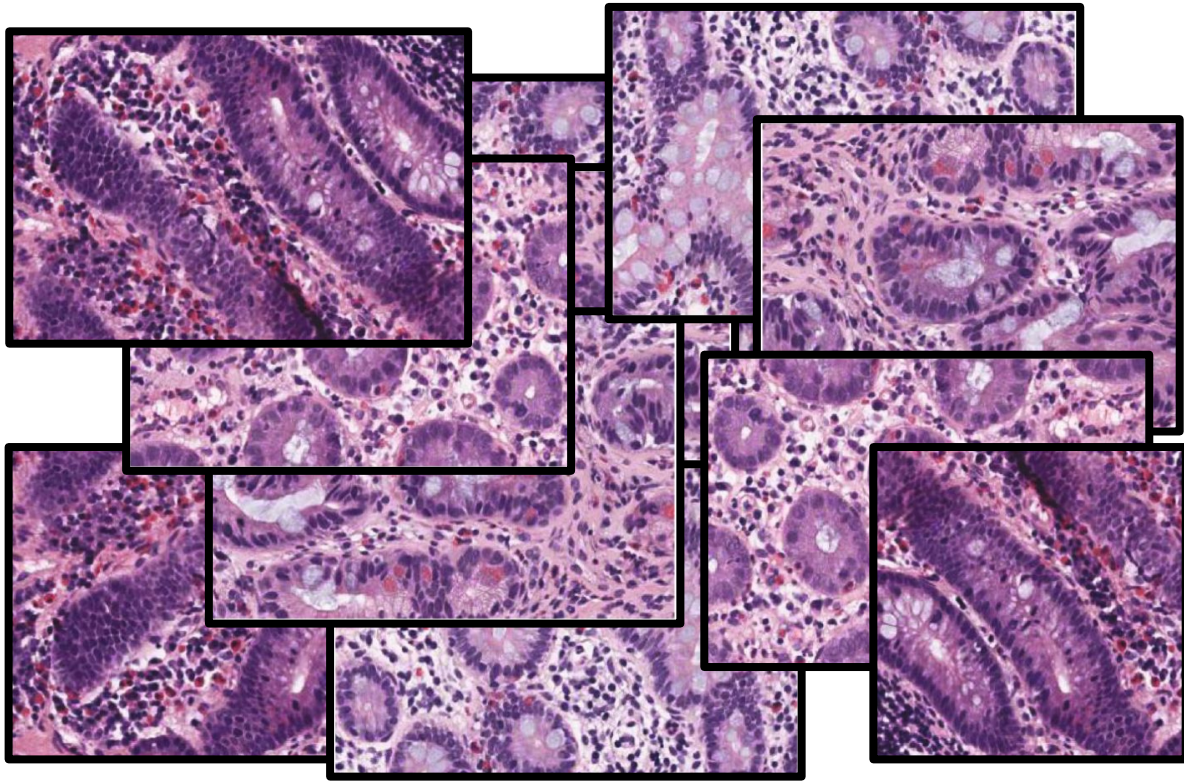


[Goldsborough, CytoGAN: Generative Modeling of Cell Images, 2017]

Semi-supervised Segmentation with GANs



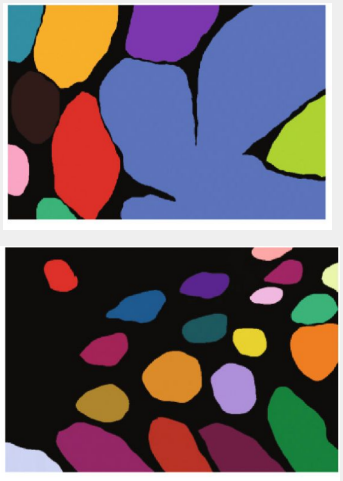
Images with segmentation labels



Images without segmentation labels

Semi-supervised Segmentation with GANs

Predicted segmentations from images that were trained on



Match distributions

Predicted segmentations from unlabelled images

Luc et al. "Semantic Segmentation using Adversarial Networks" 2016
Zhang et al., "Deep Adversarial Networks for Biomedical Image Segmentation Utilizing Unannotated Images," 2017

Semi-supervised Segmentation with GANs

Update discriminator E

$$\lambda[\underbrace{\sum_{m=1}^{M} \ell_{bce}(E(S(X_m), X_m), 1)}_{\text{E should predict 1 for labelled examples}} + \underbrace{\sum_{n=1}^{N} \ell_{bce}(E(S(U_n), U_n), 0)}_{\text{E should predict 0 for unlabelled examples}}]$$

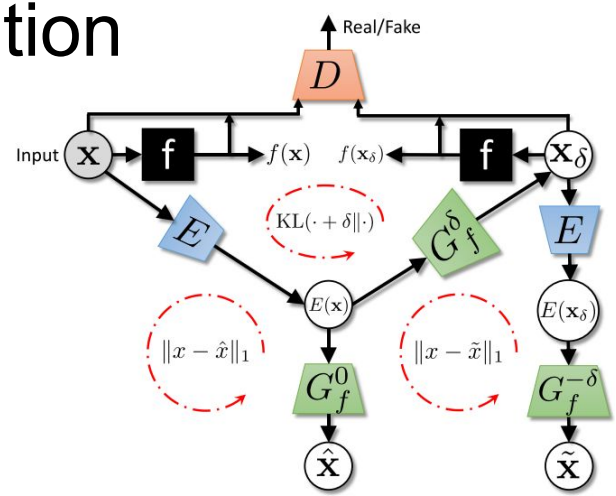
Update segmenter S

$$\underbrace{\sum_{m=1}^{M} \ell_{mce}(S(X_m), Y_m)}_{\text{Segmentation Loss}} - \lambda(\underbrace{\sum_{n=1}^{N} \ell_{bce}(E(S(U_n), U_n), 0)}_{\text{Segmentation output should not make E predict 0}})$$

Luc et al. "Semantic Segmentation using Adversarial Networks" 2016
Zhang et al., "Deep Adversarial Networks for Biomedical Image Segmentation Utilizing Unannotated Images," 2017

Explanation by Progressive Exaggeration

Train a classifier and generative model jointly while maintaining consistency between them.



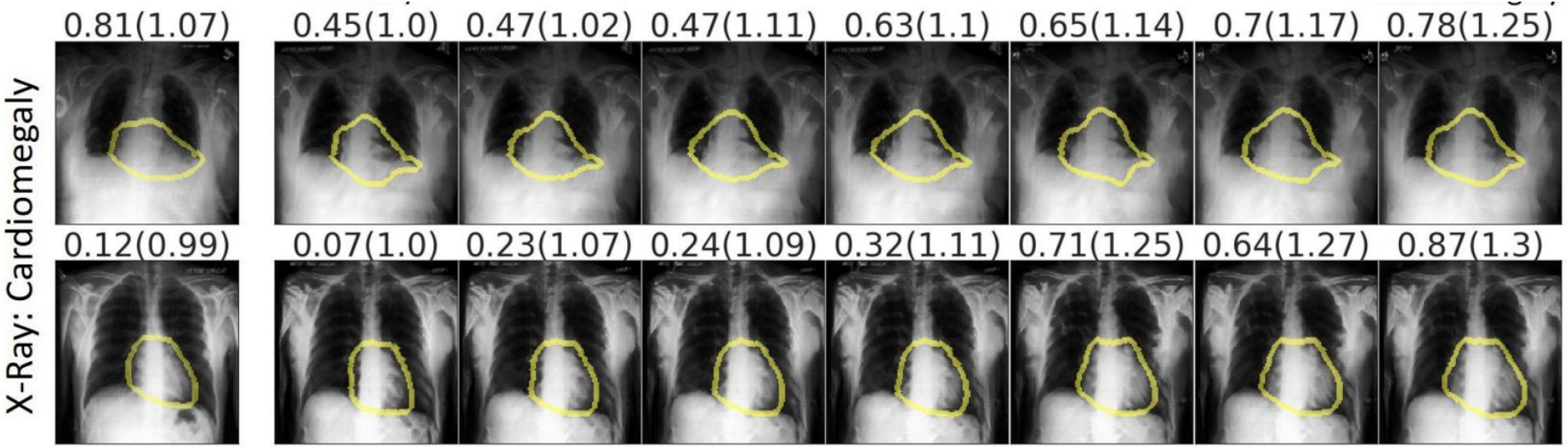
Explainer function: $\mathcal{I}_f(\mathbf{x}, \delta) = G(E(\mathbf{x}), c_f(\mathbf{x}, \delta))$

(c_f outputs a one hot)

1. **Data Consistency:** perturbed samples generated by $\mathcal{I}_f$ should lie on the data manifold, $\mathcal{M}_x$, to be consistent with real data. In other words, the generated samples should look realistic when compared to other samples.
2. **Compatibility with f:** changing the second argument in $\mathcal{I}_f(\mathbf{x}, \cdot)$ should produce the desired outcome from classifier f, *i.e.*, $f(\mathcal{I}_f(\mathbf{x}, \delta)) \approx f(\mathbf{x}) + \delta$.
3. **Self Consistency:** Applying reverse perturbation should bring $\mathbf{x}$ back to its original form *i.e.*, $\mathcal{I}_f(\mathcal{I}_f(\mathbf{x}, \delta), -\delta) = \mathbf{x}$. Also, applying setting δ to zero should return the query, *i.e.*, $\mathcal{I}_f(\mathbf{x}, 0) = \mathbf{x}$.

[Singla et al. Explanation by Progressive Exaggeration. ICLR 2020]

Explanation by Progressive Exaggeration



Generating images conditioned on an over and under prediction of the model helps explain what aspects of the image were important in prediction.
Here we can see the heart enlarge or shrink.

[Singla et al. Explanation by Progressive Exaggeration. ICLR 2020]

Chapter 5 - References

Cohen, J. P., Luck, M., & Honari, S. (2018). Distribution Matching Losses Can Hallucinate Features in Medical Image Translation. *Medical Image Computing & Computer Assisted Intervention (MICCAI)*.

Goldsborough, P., Pawlowski, N., Caicedo, J. C., Singh, S., & Carpenter, A. E. (2017). CytoGAN: Generative Modeling of Cell Images. *Workshop On Machine Learning In Computational Biology, Neural Information Processing Systems*. https://doi.org/10.1101/227645

Luc, P., Couprie, C., Chintala, S., & Verbeek, J. (n.d.). *Semantic Segmentation using Adversarial Networks*. Retrieved December 5, 2017, from https://arxiv.org/pdf/1611.08408.pdf

Zhang, Y., Lin, Y., Chen, J., Fredericksen, M., Hughes, D. P., Chen, D. Z., Yang, L., Chen, J., Fredericksen, M., Hughes, D. P., & Chen, D. Z. (2017, September 10). Deep Adversarial Networks for Biomedical Image Segmentation Utilizing Unannotated Images. *Medical Image Computing and Computer-Assisted Intervention*. https://doi.org/10.1007/978-3-319-66179-7

Singla, S., Pollack, B., Chen, J., & Batmanghelich, K. (2020, November 1). Explanation by Progressive Exaggeration. *International Conference on Learning Representations*. http://arxiv.org/abs/1911.00483